

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 19.3719 Seconds

(without alignments)
.1530.854 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGANGGALAPAAAGRL.....YSAPRDLTPPLNQTANTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	362.5	9.6	925	1 NRP2_RAT	035276 rattus norv
2	354.5	9.4	931	1 NRP2_MOUSE	035375 mus musculu
3	351.5	9.3	931	1 NRP2_HUMAN	060462 homo sapien
4	338.5	9.0	923	1 NRP1_MOUSE	097333 mus musculu
5	338	8.9	914	1 NRP1_CHICK	019795 gallus gall
6	335	8.9	923	1 NRP1_HUMAN	014786 homo sapien
7	331	8.8	922	1 NRP1_RAT	096419 rattus norv
8	331	8.8	2211	1 FAS_BOVIN	028107 bos taurus
9	330	8.8	928	1 NRP1_XENLA	028824 xenopus lae
10	317	8.4	2224	1 FAS_HUMAN	012259 homo sapien
11	314	8.3	2258	1 FAS_PIG	099181 sus scrofa
12	299	7.9	387	1 MFGM_HUMAN	008431 mus musculu
13	297.5	7.9	427	1 MFGM_RAT	070480 rattus norv
14	285.5	7.6	2351	1 FAS_HUMAN	000451 homo sapien
15	276.5	7.3	463	1 MFGM_MOUSE	021936 mus musculu
16	276.5	7.3	2319	1 FAS_MOUSE	006194 mus musculu
17	271.5	7.2	2133	1 FAS_PIG	012283 sus scrofa
18	268	7.1	427	1 MFGM_BOVIN	095114 bos taurus
19	259.5	6.9	409	1 MFGM_PIG	079385 sus scrofa
20	229.5	6.1	280	1 XLR1_FUGRO	096675 tuugu rudrip
21	221.5	5.9	224	1 XLR1_MOUSE	092114 mus musculu
22	218.5	5.8	224	1 XLR1_HUMAN	015537 homo sapien
23	199	5.3	1308	1 CTAA_HUMAN	090840 homo sapien
24	195.5	5.2	1331	1 CTAA_HUMAN	090840 homo sapien
25	194.5	5.2	1310	1 CTAA_MOUSE	090840 mus musculu
26	192.5	5.1	1284	1 NRP4_DROME	094887 drosophila
27	190	5.0	3133	1 HMC2_BOVINO	098092 bombyx mori
28	186.5	4.9	686	1 HMC2_HUMAN	000187 homo sapien
29	186.5	4.9	1288	1 CTAS2_HUMAN	092876 homo sapien
30	172.5	4.6	514	1 CTAS2_XENLA	042664 xenopus lae
31	165	4.4	986	1 BMP1_HUMAN	013497 homo sapien
32	163	4.3	408	1 PCO1_RAT	008628 rattus norv
33	162.5	4.3	704	1 PCO1_MOUSE	098064 mus musculu

34	162.5	4.3	854	1 DDR2_MOUSE	062371 mus musculu
35	161.5	4.3	449	1 PCO1_HUMAN	015113 homo sapien
36	160	4.2	468	1 PCO1_MOUSE	061398 mus musculu
37	159	4.2	550	1 COCH_HUMAN	043405 homo sapien
38	158	4.2	991	1 BMP1_MOUSE	098063 mus musculu
39	158	4.2	1022	1 TLID_BRARE	057460 brachydanio
40	157	4.2	552	1 COCH_MOUSE	062507 mus musculu
41	156	4.1	707	1 BMP1_XENLA	098070 xenopus lae
42	155.5	4.1	855	1 DDR2_HUMAN	016832 homo sapien
43	149.5	4.0	695	1 CASP_MESAU	015156 mesocricetu
44	142.5	3.8	1385	1 CTAL_MOUSE	054991 mus musculu
45	142	3.8	855	1 ST14_HUMAN	097576 homo sapien

ALIGNMENTS

RESULT 1	ID	NRP2_RAT	STANDARD	PRT	925 AA.
AC	035276				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).				
GN	NRP2.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
ON	NCBI_TaxID=10116;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Sprague-Dawley;				
RX	MEDLINE=97433085; PubMed=9288754;				
RA	Kotodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,				
RA	Ginty D.D.;				
RT	"Neuropilin is a semaphorin III receptor.";				
RL	Cell 90:753-762(1997).				
CC	- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165				
CC	- AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.				
CC	- SUBCELLULAR LOCATION: TYPE I membrane protein.				
CC	- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE				
CC	CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE				
CC	LINING IN THE RIBS.				
CC	- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.				
CC	- SIMILARITY: CONTAINS 2 CUB DOMAINS.				
CC	- SIMILARITY: CONTAINS 2 F3/8 TYPE C DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 MAM DOMAIN.				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION				
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION				
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS				
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY				
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL				
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/				
CC	OR SEND AN EMAIL TO license@sib-sib.ch).				
DR	EMBL: AF016297; AAC53338.1; -				
DR	HSSP: P12259; IC2T.				
DR	InterPro: IPR000859; CUB_domain.				
DR	InterPro: IPR000421; FAS5_C.				
DR	InterPro: IPR000998; MAM_domain.				
DR	Pfam: PF00431; CUB: 2.				
DR	Pfam: PF00629; MAM: 1.				
DR	Pfam: PF00754; F5_F8_Type_C: 2.				
DR	SMART: SM00242; CUB: 2.				
DR	SMART: SM00231; FAS5C: 2.				
DR	SMART: SM00137; MAM: 1.				
DR	PROSITE: PS01180; CUB: 2.				
DR	PROSITE: PS01285; FAS5C_1: 2.				
DR	PROSITE: PS01286; FAS5C_2: 2.				
DR	PROSITE: PS50060; MAM_2: 1.				

Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.

KW SIGNAL 1 22
 FT CHAIN 23 925
 FT DOMAIN 23 858
 FT TRANSMEM 859 883
 FT DOMAIN 884 925
 FT DOMAIN 28 142
 FT DOMAIN 149 267
 FT DOMAIN 277 427
 FT DOMAIN 434 592
 FT DOMAIN 642 802
 FT DISULFID 83 105
 FT DISULFID 149 175
 FT DISULFID 208 230
 FT DISULFID 277 427
 FT DISULFID 434 592
 FT CARBOHYD 152 157
 FT CARBOHYD 157 157
 FT CARBOHYD 629 629
 FT CARBOHYD 833 833
 FT CARBOHYD 834 834
 SQ SEQUENCE 925 AA; 103896 MW; 3BF62903F644851C CRC64;

Query Match 9.6%; Score 362.5; DB 1; Length 925.
 Best Local Similarity 22.3%; Pred. NO. 7.2e-18;
 Matches 143; Conservative 100; Mismatches 246; Indels 151; Gaps 24;

41 CGHLYTODSGTMTSKNPGTYPNHTVEKTYPRK-GKRLIRLG-DIDIESQTCASDY 98
 28 CGGRINSDAGYITSPGYRQDYPHONCEWVYAPENPKIYLNPNHEIKHCKYDF 87
 99 LIF-----TSSDQYGPYCGSMTPKELLINSEVTFRESGHISGRGLTY-----AS 149
 88 IEIRDGESEADLKHGCGNI-APPTIISGCVLYIKFSDYAROGAGFSLREIFKTS 146
 150 SD-----HPLILIC-----LBRASHYLTKEYSK----- 173
 147 EDCSNFTSPNGTISPGEPPEKYPHNLCTFTILAKPMEIILGLFTDLEHDPLOYGEG 206
 174 -CPACGRVADGIS--GNMVDGYRDTSLCKNAIHAGIADLGGQISVLRKGISRY-- 228
 207 DCKYMWLDIMGIPHVGGLICKYCTKPKSKRSSTGILSLFPHDMVAVKCFARYYL 266
 229 -----EGLANGVLSRD-GSLSDKRFLETSNGCSRSLSFEPDGOIRASSSMQVNESGD 281
 267 VHOEPPEFQCAAPLGMESGRIANE-----QISASSTR-----S 300
 282 QVHNSPGARLDODGSPNASGSSNNHNPRLLEIDLGKKKITGIRTTG--STQSNFNF 339
 301 DGRWTPQOSRLHGDNGMTPNVDSN-----KEYLOYDLNRLMFLMTALATOGAISRETQKGY 356
 340 YKSFVNFNFKNNSKMYKTYGIVNNEKVFQNSNFRDPVONNETPTPYARVYVQPM 399
 357 YKSYKLELVSTNGEDMWMYRHKGNH--KYFQANNDATELVYLNKKLPTLRIRIRPQM 414
 400 HORLAKVELLGCQITQGNDSLVMKTS-----OSTSVSTKKEDETTTPT----- 445
 415 HLGIALRLLELGCRCVTDAPCSNMLGLJADTQISASTREYIMSPSAARLVSSRSWG 474
 446 -----PSEETSGINITTVAIPVLVYLVFAGMGIFA-----AFPKKKKKSGPYSA 493
 475 FPRNPOAPGEEW-LQVLDLGTPTKYKGVITQARGSGDITLAEARAFYKKEFVSYSLNKG 533
 494 EAOKTDCWKQIKYFPAHQSA-EFTISYD-----NEKEMTOKL 530
 534 D-----WEYIODPTQOPKLEEGNMHMDTPDIRFEPVPAQYVYVPERMSPACIGRL 587
 531 DLITSDMADYQOPL-MIGTGIVTRKGSFFRPMDDAEAG 569
 588 EYLGCDMTDSKPTVETLGPVAKSEETTPYPMDDEATECG 627

RESULT 2
 NR2_MOUSE
 ID NR2_MOUSE STANDARD: PRI: 931 AA.
 AC 035375; 035373; 035374; 035376; 035377; 035378;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-UN-2002 (Rel. 41, Last annotation update)
 DE Neutropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
 GN NR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RP STRAIN-BALB/c;
 RX MEDLINE=97470888; PubMed=9331348;
 RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
 RT "Neutropilin-2, a novel member of the neutropilin family, is a high affinity receptor for the semaphorins Sema E and Sema IV but not Sema III.";
 RT III.";
 RL Neuron 19:547-559(1997).
 CC - FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLEC-2 ISOFORM OF PEG.
 CC - SUBUNIT: NEUTROPHILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH NEUTROPHILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - ALTERNATIVE PRODUCTS: 6 ISOFORMS; A22 (SHOWN HERE), A0, A5, A17, B0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES, INTESTINAL EPITHELIUM, KIDNEY, LONG AND SUBMANDIBULAR GLAND.
 CC - DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND IS DEVELOPMENTALLY REGULATED.
 CC - SIMILARITY: BELONGS TO THE NEUTROPHILIN FAMILY.
 CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC - SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC - THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF022856; AAC53379.1; -
 DR EMBL; AF022854; AAC53377.1; -
 DR EMBL; AF022855; AAC53378.1; -
 DR EMBL; AF022857; AAC53380.1; -
 DR EMBL; AF022858; AAC53381.1; -
 DR EMBL; AF022861; AAC53382.1; -
 DR HSSP; P12259; ICRF.
 DR MGD; MGI:1100492; NRP2.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS50060; MAM_2; 1.
 DR Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 931
 FT POTENTIAL.
 FT NEUTROPHILIN-2.

DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00231; FA58C; 2.
 DR SMART: SM00137; MAM; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01285; FA58C; 1; 2.
 DR PROSITE: PS01286; FA58C; 2; 2.
 DR PROSITE: PS00740; MAM; 1; 1.
 DR PROSITE: PS50060; MAM; 2; 1.
 KM Transmembrane: Glycoprotein; Neurone; Signal; Repeat; Receptor.
 FT SIGNAL 1 21
 FT CHAIN 1 923
 FT DOMAIN 22 856
 FT TRANSMEM 857 879
 FT DOMAIN 880 923
 FT DOMAIN 27 141
 FT DOMAIN 147 265
 FT DOMAIN 275 424
 FT DOMAIN 431 583
 FT DOMAIN 645 811
 FT DISULFID 27 54
 FT DISULFID 82 104
 FT DISULFID 147 173
 FT DISULFID 206 228
 FT DISULFID 275 424
 FT DISULFID 431 583
 FT CARBOHYD 150 261
 FT CARBOHYD 261 300
 FT CARBOHYD 300 322
 FT CARBOHYD 522 542
 FT CARBOHYD 842 842
 SQ SEQUENCE 923 AA; 103020 MW; 06448BA170796808 CRC64;

Query Match 9.0%; Score 338.5; DB 1; Length 923;
 Best Local Similarity 20.0%; Pred. No. 3.6e-16;

Matches 158; Conservative 106; Mismatches 267; Indels 227; Gaps 30;

DR 17 RGL-----LALLAVSAPLRLQAEELGDCGHLVYODSGTMSKNRYPTGYNHTYCEK 70
 DB 3 RGLPLCLATLALALAGNR-----SDKCGTITKIENGNYLTSTGYRHSYSEKCEW 56
 QY 71 TITVPGK-KRLIIRLG-DLIDIESQTCASDYLFTSSD---YGPYCGSMVTPKELLN 124
 DB 57 LIOAPEPYRRIITINFPNPHDLERDCKYVEYIDENEGRLMKFCGKI-APSPVSS 115
 QY 125 TSEVTVRPFGSHISGRFLTYA-----SSDHDLLTTC-- 158
 DB 116 GPELFTKEVSDYETGAGFSIRIETIKRQDECSQNTAFTGVIKSPGPEKRYNCLCTY 175
 QY 159 -----TERASHYKTEYSKPCPAG--CRVAGDISGNMVDGYRDTSLCKRAIH 205
 DB 176 IIFAPKSEILIEFESEFDLEQDSNP--PGMPCRYDRLEI---WGCFPE-----YGRH 223
 QY 206 AG-TIDELAGGI-----SVLQKRGISRYEGITLNGVLSRDGSLSDRFLTE 250
 DB 224 IGYNCCQKPRGRIRSSSGVLSWYFYTDSAIKAGFSANTSVLQSSI-SEDFK----- 274
 QY 251 TSNCGSRISLSEF-----DQIRASSSMOSVNEGSDQVHNSPGARLQDGPMSASDSSN 306
 DB 275 -----CHEALGMSGEHSDQITNSQYGT-----NNSVESRSLNTPENGMTPEEDST 322
 QY 307 NHRPREMLIDLGKKKRTIGIRTTG--STQSNFNPFYKSFVNFKNNNSMKTYKGIYNN 364
 DB 323 -----KMIQVDDGLRFAVAVGTGCAISKEKKYVYKRYRDISNGEDMISLK--EGN 376
 QY 365 EEKVFQGNSTNRPDVPNNFTIPPVARYRVVYVOTWQRIALAKVELIGCOITQGNDSLWVR 424
 DB 377 KAIIFQGNSTNRPDVPVYGVFSKPLITFRVRIKIPVSMETGISMREYVGCKITD-----YF 430
 QY 425 KTSQSTSVSGKKDEETITPPISEETSTGINITY-----AIPVLVAVLVAGGIR 478
 DB 431 CSGMLGMSGLISDSQITNSQADNRMMFENILVTSRTGMLP----- 474

QY 479 AFRKKKKSGPYGSAENAKTDCWKQI-----KYPPARHOSAEFTISYDN 522
 DB 475 -----PSRPHY-----TNEMLQVLDGEEKIVRGVYIIOGCHRRNKRYMKRFAIYSN 521
 QY 523 EKEMTOKLIDITSDMADYQOPLMIGTVTRKSTFRPMDTDAEAGVSTADGHHYDCPO 582
 DB 522 -----NSDMWTIMDDSKRRKASPEGNNNTDTP 550
 QY 583 RARHHEALPLAPEPEYATPI---VERHVLKRAHTSASGIVRPPQPHKHSLSGGR 639
 DB 551 LRTFSPSLSTFIRIYPERAHSGLRLMELGCEVAPGAPTPPNPNVHE----- 602
 QY 640 SPVAGVAGODGYRPHSAPADRGYDRKAVSALATE 677
 DB 603 -----CDDQDANCHSGIGD--FOLGGTIVLATE 630

RESULT 5
 NRPL_CHICK
 ID NRPL_CHICK STANDARD; PRT; 914 AA.
 AC P19795;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilin-1 precursor (A5 protein).
 GN NRPL OR NRP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White Leghorn; TISSUE=Embryonic brain;
 RX MEDLINE=95324761; PubMed=7601310;
 RA Takagi S., Kasuya Y., Shimizu M., Matsuura T., Tsuboi M., Kawakami A.,
 RA Fujisawa H.;
 RT Expression of a cell adhesion molecule, neurofilin, in the
 RT developing chick nervous system.";
 RL Dev. Biol. 170:207-222(1995).
 CC - FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANISMS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
 CC PROPERTIES.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
 CC (LAYERS D AND E OF SGEs), AMACINE CELLS OF RETINA, NEURITES OF
 CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
 CC BLOOD VESSELS IN THE ENTIRE EMBRYO.
 CC - SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC - SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D45416; BAA08256.1; -
 CC HSSP: P12259; IC2F.
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR000421; FA58_C.
 CC InterPro: IPR000998; MAM_domain.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF00629; MAM; 1.
 CC Pfam: PF00754; F5_F8_type_C; 2.

DR PRINTS; PRO0020; MAMDOMAIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00060; MAM_2; 1.
 DR Transmembrane; Glycoprotein; Neurone; signal; Repeat; Receptor;
 K Cell adhesion.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 914 NEUROPILIN-1.
 FT DOMAIN 20 847 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 848 870 POTENTIAL.
 FT DOMAIN 871 914 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 25 139 CUB 1.
 FT DOMAIN 145 263 CUB 2.
 FT DOMAIN 273 422 F5/8 TYPE C 1.
 FT DOMAIN 429 581 F5/8 TYPE C 2.
 FT DOMAIN 636 801 MAM.
 FT DISULFID 25 52 PROBABLE.
 FT DISULFID 80 102 PROBABLE.
 FT DISULFID 145 171 PROBABLE.
 FT DISULFID 204 226 PROBABLE.
 FT DISULFID 273 422 BY SIMILARITY.
 FT DISULFID 429 581 BY SIMILARITY.
 SQ SEQUENCE 914 AA; 102480 MW; DD2EE6DF0CB868C CRC64;

Query Match Best Local Similarity 25.78; Pred. No. 3.9e-16; Score 338; DB 1; Length 914;

Matches 115; Conservative 67; Mismatches 160; Indels 86; Gaps 17;

QY 21 ALLAVSAPRLQAELEDCGHLVYDSCGTMSTSKNPGTYPMNTVCETKTYPKG-KR 79
 11 ALFTLSALR-----SDKGDRIKILSPGYLSPGPGYHSHQKEMILQAEPRQR 64
 QY 80 LILKLG-DLIESOTCASDYLLFTSSSQ-----YGYGSGSTVPEKILLNTSEYTPRES 134
 65 IMINPNPHFDEDRDCKYDVEYIDGNABEGRMLGKCYCKI-APPPIVSGPYLFKFEVS 123
 QY 135 GSHISGRFELTY-----ASS-----DHPDLITG----- 158
 124 DYEHGAGFSRIYEFKKGPECSRNFTSSSGMITSPPGPEKYPNSLETTYIIPAKSEI 183
 QY 159 -LERASHYLKTEYSKFCPCAGCRVADGISGMVGYRDT-----SLCKRAIHAGIADDEL 213
 184 ILERSEFLEPDSNTPGGAFCRDRLEI-----WDGFPVGHIGRYC-GQNNPGVRST 238
 QY 214 GGOISVLRKGISREGLANGVLSRDSLSDKRFLFTSNGCSNLSLSEF-----DGOIRA 269
 239 GILSMVEYTDISAIRKEGFSANYSVSSVSSEDFO-----CMELGMESEIHSDQITV 291
 QY 270 SSSMOSVSGSDQYHMSGGLADQGPVSGDSSNNHKKREMLDGLGKKKITGIRT 329
 292 SSOYSAT-----WSEKSRLLTPENGWTPGEOS-----VREMLQVDLGLLRVYSGIGT 339
 QY 330 TG--STOSNFNFYKSFYPMNFKNNNSKKYKGIIVNNEKVFQGSNFRDPVQNNFTPIPI 387
 340 OGALSKREKKREYLYKTYRVDVDSNGEDMITLK--EGNNPVYFGQNSNPTDYYVYRPPRPV 397
 QY 388 VARYRVVVPQTHORIALKVELIGCOIT 415
 398 LTRVRIKIPVSMENGVSILREYVCKIT 425
 DB 398 LTRVRIKIPVSMENGVSILREYVCKIT 425

RESULT 6

NRPI_HUMAN STANDARD; PRT; 923 AA.
 AC 014786; 060461;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165
 DE receptor).
 GN NRPI OR NRP OR VEGF165R.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
 RX MEDLINE=97433084; PubMed=9288753;
 RA He Z., Tessier-Lavigne M.;
 RT "Neuropilin is a receptor for the axonal chemorepellent semaphorin
 IT Cell 90:739-751(1997).
 (2)
 SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
 RP TISSUE-Breast;
 RX MEDLINE=98188099; PubMed=9529250;
 RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
 RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
 isoform-specific receptor for vascular endothelial growth factor";
 RL Cell 92:735-745(1998).
 (3)
 SEQUENCE FROM N.A. (SOLUBLE/SNRP1 ISOFORM), AND SEQUENCE OF 22-31.
 RP TISSUE-Prostatic adenocarcinoma;
 RX MEDLINE=20183929; PubMed=10688880;
 RA Gagnon M.L., Bieleberg D.R., Gechtman Z., Miao H.-Q., Takashima S.,
 RA Soker S., Klagsbrun M.;
 RT "Identification of a natural soluble neuropilin-1 that binds vascular
 endothelial growth factor: In vivo expression and antitumor
 RT activity";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
 (4)
 CHARACTERIZATION.
 RX MEDLINE=20309748; PubMed=10748121;
 RA Guzman-Pollock Z., Cohen T., Herzog Y., Neufeld G.;
 RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
 form of vascular endothelial growth factor (VEGF) and of placenta
 RT growth factor-2, but only neuropilin-2 functions as a receptor for
 RL the 145-amino acid form of VEGF";
 RL J. Biol. Chem. 275:18040-18045(2000).
 (1)
 FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE
 DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE
 FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS
 OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT
 ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2
 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B.
 COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR
 AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED
 ANGIOGENESIS.
 (2)
 FUNCTION: THE SOLUBLE/SNRP1 ISOFORM BINDS VEGF-165 AND APPEARS
 TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY
 SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE
 SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD
 VESSEL NUMBER AND INTEGRITY.
 (3)
 SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. THE SOLUBLE/SNRP1
 ISOFORM IS SECRETED.
 (4)
 ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEMBRANE-BOUND (SHOWN HERE)
 AND SOLUBLE/SNRP1; ARE PRODUCED BY ALTERNATIVE SPLICING.
 (5)
 TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND
 SOLUBLE/SNRP1 ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-
 BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT
 TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE
 NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART
 AND PLACENTA; MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY
 AND PANCREAS; AND LOW IN ADULT BRAIN. THE SOLUBLE/SNRP1 ISOFORM IS
 FOUND IN LIVER HEPATOCYTES, KIDNEY DISTAL AND PROXIMAL TUBULES.
 (6)
 SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 (7)
 SIMILARITY: CONTAINS 2 CUB DOMAINS.
 (8)
 SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 (9)
 SIMILARITY: CONTAINS 1 MAM DOMAIN.
 (10)
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF018956; AAC51759.1; -
DR EMBL: AF016050; AAC12921.1; -
DR EMBL: AF145712; AAF4344.1; -
DR HSSP: P12259; 1C2T.
DR HSSP: HGNC:8004; NRPI.
DR MIM: 602069; -
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR001092; HLH_basic.
DR InterPro: IPR000998; MAM_domain.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM0137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00760; MAM_2; 1.
DR Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
KW -----
FT CHAIN 1 21
FT SIGNAL 1 21
FT DOMAIN 22 923 NEUROPIILIN-1.
FT TRASMEM 857 879 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 880 923 POTENTIAL.
FT DOMAIN 880 923 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 141 CUB 1.
FT DOMAIN 147 265 CUB 2.
FT DOMAIN 275 424 F5/8 TYPE C 1.
FT DOMAIN 431 583 F5/8 TYPE C 2.
FT DOMAIN 645 811 MAM.
FT DISULFID 27 54 PROBABLE.
FT DISULFID 82 104 PROBABLE.
FT DISULFID 147 173 PROBABLE.
FT DISULFID 206 228 PROBABLE.
FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 583 BY SIMILARITY.
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 645 644 MISSING (IN SOLUBLE/SNRPI ISOFORM).
FT VARSPLIC 645 923 K -> E (IN REF. 1).
FT CONFLICT 26 26 D -> H (IN REF. 2).
FT CONFLICT 749 749 E -> D (IN REF. 2).
FT CONFLICT 855 855 E -> D (IN REF. 2).
SQ SEQUENCE 923 AA: 103120 MM: 103120 MD: ADEADCA48A9E5D57 CRC64:
Query Match 8.9%; Score 335; DB 1; Length 923;
Best Local Similarity 24.9%; Pred. No. 6,4e-16;
Matches 124; Conservative: 76; Mismatches 176; Indels 122; Gaps 23;

DB 116 GPELF1KFSVDYETHGAGSIRYEIPKRGPCSQNTTPSGVILKSPGPEKYPNSLECTY 175
QY 159 -----LERASHYKTERXSKFCPAG--CRDVAIGISGMVGYDTSILCAAH 205
DB 176 IYAPAKSEILLEFESEFDLEPDSNP--PGMFCRYRLEI---WDGFD-----VGRH 223
QY 206 AG-TIADLEGQI-----SVLQRKISRVEGIIANGVLSRDGSLDRKFLP 250
DB 224 IGRYCGQKTPRGRIRSSSGILSMVFYTDIAIAKGFANISVLAGSSV-SEDFK----- 274
QY 251 TSNCCSRSLSEFP---DDQIRASSSMOSVNESGDVHMSPGARLQDDGPPSAISDSSN 306
DB 275 ----CMEALGMESEGIHSDQITASSQYST-----NMSAERSRLNYPENGWTPCEDSY 322
QY 307 NHPREMLEIDGEEKKRTGIRPTG--SQOSNENFVYKSPVMEKNNNSKMKTKGIYVN 364
DB 323 ----REMIOVDGLLFYAVAGTQGISKTKKTKYKTKIDVSSNGEDWITIK--EEN 376
QY 365 EEKVFGNSNFRDPVONNIPPIVARYVVPOTWRIALRYELIGCOIT-----OGNDS 420
DB 377 KPYLFGQNTNPDTVVAVPFPKPLTFRFAIKPATWETGISMREYVCGKITDPCSGMLG 436
QY 421 LVWRTKSQSTSVSTKED 438
DB 437 MVSGLISDQITSSNOGD 454

RESULT 7
NRPI_RAT
ID NRPI_RAT STANDARD; PRT; 922 AA.
AC 090MD9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurophilin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
GN NRPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=97433085; PubMed=9288754;
RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J., Ginty D.D.
RT Neurophilin is a semaphorin III receptor.*;
RL Cell 90:753-762(1997).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOREGULATION OF SEMAPHORINS. IT
CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
CC BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGE, THE VEGF-165
CC ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
CC IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: FOUND IN THE EMBRYONIC NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE NEUROPHILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF016296; AAC53337.1; -

FT DOMAIN 348 686 F5/8 TYPE A 2.
 FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 535 686 PLASTOCYANIN-LIKE 4.
 FT SIMILAR 696 1564 B.
 FT DOMAIN 899 915 TO 17 AA REPEATS IN HUMAN FAS.
 FT DOMAIN 1124 1151 2 X 14 AA TANDEN REPEATS.
 FT REPEAT 1124 1137 1-1.
 FT REPEAT 1124 1137 1-2.
 FT REPEAT 1188 1453 30 X 9 AA APPROXIMATE TANDEN REPEATS OF
 [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
 FT REPEAT 1188 1196 2-1.
 FT REPEAT 1197 1205 2-2.
 FT REPEAT 1206 1214 2-3.
 FT REPEAT 1215 1223 2-4.
 FT REPEAT 1224 1232 2-5.
 FT REPEAT 1233 1241 2-6.
 FT REPEAT 1242 1250 2-7.
 FT REPEAT 1251 1259 2-8.
 FT REPEAT 1260 1268 2-9.
 FT REPEAT 1269 1277 2-10.
 FT REPEAT 1278 1286 2-11.
 FT REPEAT 1287 1295 2-12.
 FT REPEAT 1296 1304 2-13.
 FT REPEAT 1305 1313 2-14.
 FT REPEAT 1314 1322 2-15.
 FT REPEAT 1323 1331 2-16.
 FT REPEAT 1332 1340 2-17.
 FT REPEAT 1341 1349 2-18.
 FT REPEAT 1350 1358 2-19.
 FT REPEAT 1359 1367 2-20.
 FT REPEAT 1368 1376 2-21.
 FT REPEAT 1377 1385 2-22.
 FT REPEAT 1386 1394 2-23.
 FT REPEAT 1395 1403 2-24.
 FT REPEAT 1404 1412 2-25.
 FT REPEAT 1413 1421 2-26.
 FT REPEAT 1422 1430 2-27.
 FT REPEAT 1431 1439 2-28.
 FT REPEAT 1440 1444 2-29 (PARTIAL).
 FT REPEAT 1445 1453 2-30.
 FT DOMAIN 1569 1890 F5/8 TYPE A 3.
 FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 1894 2048 F5/8 TYPE C 1.
 FT DOMAIN 2053 2208 F5/8 TYPE C 2.
 FT SITE 741 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT DISULFID 167 193 PROBABLE.
 FT DISULFID 499 525 PROBABLE.
 FT DISULFID 1712 1738 BY SIMILARITY.
 FT DISULFID 1894 2048 BY SIMILARITY.
 FT DISULFID 2053 2208 BY SIMILARITY.
 FT MOD_RES 697 701 SULFATION (POTENTIAL).
 FT MOD_RES 701 701 SULFATION (POTENTIAL).
 FT MOD_RES 730 730 SULFATION (POTENTIAL).
 FT MOD_RES 1513 1513 SULFATION (POTENTIAL).
 FT MOD_RES 1529 1529 SULFATION (POTENTIAL).
 FT MOD_RES 1537 1537 SULFATION (POTENTIAL).
 FT MOD_RES 1541 1541 SULFATION (POTENTIAL).
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 587 592 N-TERMINAL (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;
 Query Match 8.8%; Score 331; DB 1; Length 2211;
 Best Local Similarity 32.5%; Pred. No. 4,4e-15;
 Matches 88; Conservative 48; Mismatches 89; Indels 46; Gaps 10;
 QY 162 ASHTKTEY-SKCPA-----GCRVAGDISGNV--DGYRDSILCKAIIHGIIDE 212
 DB 1968 AKHTKPYTTEFCVAYSLDRKNRIFKGNSTRVMTFGNSDASTIKENDIPPVVA-- 2025
 QY 213 LGQISVLRKRGISREGLIANGVLSRDSLSDRKFLFTSGCSRSISFE---PDGOIR 268
 DB 2026 -----RYRISPTGSKYKRALDLQGC-EVAGCSTPLMGEGKLEKQIT 2070
 QY 269 ASS---SWQSVNESGDQVHWSFGQARLDQD--PSMASGSDSSNNHKKPREMLDILGKKR 323
 DB 2071 ASSEKSKSWWG-----NYMEPLRLAAGRVANAKANNNN---QTLQIDLKIKK 2119
 QY 324 ITGRTGTSQSNFNFYKSVNFKNNNSKWKYTKGIVNNEEVPOGNSFRPVOYNNF 383
 DB 2120 ITAIVTQCKLSSEMYKSTYTHSDGTMDKPRERKSSWVDKIFGNNNVKGVKNFF 2179
 QY 384 IPPVARYRVVPOTMORIALKELIGCOI 414
 DB 2180 NPILSRIRIRIPKTNWSIALRELPGCDM 2210
 RESULT 9
 NRPL_XENIA STANDARD; PTR; 928 AA.
 AC P28824;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein) (A5 antigen).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Piploidea;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91337456; PubMed=1908252;
 RA Takagi S., Hirata T., Agata K., Mochi M., Eguchi G., Fujisawa H.;
 RT "The A5 antigen, a candidate for the neuronal recognition molecule,
 has homologues to complement components and coagulation factors.";
 RL Neuron 7:295-307(1991).
 CC - FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE OF THE NERVOUS
 SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
 BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
 CC - SURCELLULAR LOCATION: TYPE I membrane protein.
 CC - TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
 CC NEURONS.
 CC - SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.


```

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D10467; BAA01260.1; .
CC HSSP; P12259; 1C2T.
CC InterPro: IPR000859; CUB_domain.
CC InterPro: IPR000421; FA58_C.
CC InterPro: IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00231; FA58C; 2.
CC SMART; SM00137; MAM; 1.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS0060; MAM_2; 1.
CC Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Antigen.
KM SIGNAL.
FT 1 21 POTENTIAL.
FT 2 928 NEUROFILIN-1.
FT 22 860 EXTRACELLULAR (POTENTIAL).
FT 22 860 TRANSMEM.
FT 861 883 POTENTIAL.
FT 884 928 CYTOPLASMIC (POTENTIAL).
FT 27 141 CUB 1.
FT 27 141 CUB 2.
FT 265 F5/8 TYPE C 1.
FT 275 424 F5/8 TYPE C 2.
FT 431 584 MAM.
FT 431 584 MAM.
FT 466 812 PROBABLE.
FT 27 54 PROBABLE.
FT 82 104 PROBABLE.
FT 147 173 PROBABLE.
FT 206 228 PROBABLE.
FT 275 424 BY SIMILARITY.
FT 431 584 BY SIMILARITY.
FT 431 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 928 AA; 103416 MW; AF68323B0AAC789D CRC64;

Query Match 8.8%; Score 330; DB 1; Length 928;
Best Local Similarity 24.2%; Pred. NO. 1.5e-15;
Matches 112; Conservative 74; Mismatches 17; Indels 106; Gaps 20;

```

```

OY 213 LGGQISVLQKRGISRYEGILA-----NGVLSRDG-----SLSDKRFLETSNGCSRS 258
DB 224 IGRYCGMTPGVRNRFETILSMIFHTDAIAKEGFANFSVQSTFDDF-----QCKEA 278
OY 259 LSEFPD---GOIRASSWSQVNSGSDVHMSPGARLDQGPSASDSSNNKPREML 314
DB 279 LGMESGEIHFDDISVSQY-----SMNMSAERSRLNLYVENCWTPGEDT-----VKEMI 326
OY 315 EIDLKERRKIGIRFTG---STQSNFNFYKSPVMAFKNNNSWKYKGIIVNNEEKVPOGN 372
DB 327 QVDLENLRFVSGITQGAISKEFKKTKYKVDISSNGEDWTLKO--GNKHILVFEN 384
OY 373 SNFRDPVQNNFIPRYAVRVVVPOTWQRIALKEVILCCQIT 415
DB 385 TDATDVYVRPFESKPYITRFVRLRPFTWENGISLRELIGCKRT 427

RESULT 10
FAS_HUMAN STANDARD; PRT; 2224 AA.
ID FAS_HUMAN
AC P12259; O14285;
DT 01-OCT-1989 (Rel. 12, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92232668; PubMed=1567832;
RA Crilpe L.D., Moore K.D., Kane W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL Biochemistry 31:3777-3785(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87260886; PubMed=3110773;
RA Jenny R.J., Pittman D.D., Toole J.J., Kitz R.W., Aldape R.A.,
RA Hewick R.M., Kaufman R.J., Mann K.G.;
RT "Complete cDNA and derived amino acid sequence of human factor V.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
RN [3]
RP SEQUENCE OF 1-1600 FROM N.A.
RX MEDLINE=88107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
RT of internal repeats.";
RL Biochemistry 26:6508-6514(1987).
RN [4]
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE=86313665; PubMed=3092220;
RA Kane W.H., Davie E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
RT factor homologous to factor VIII and ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=93203619; PubMed=8454869;
RA Shen N.L., Fan S.-T., Pyatt J., Graff R., Lapolla R.J.,
RA Edgington T.S.;
RT "The serine protease cofactor factor V is synthesized by
RT lymphocytes.";
RL J. Immunol. 150:2992-3001(1993).
RN [6]
RP SULEFATION.
RX MEDLINE=94264012; PubMed=8204629;
RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
RA Kaufman R.J.;

```

RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity."
 RL Biochemistry 33:6952-6959(1994).
 RP [17]
 RN SULFATION.
 RX MEDLINE-90366699; PubMed-2168225;
 RA Horton G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V."
 RL Blood 76:946-952(1990).
 RN [18]
 RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RP MEDLINE-20052169; PubMed-10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V."
 RL Nature 402:434-439(1999).
 RN [19]
 RX VARIANT APCR GLN-534.
 RP MEDLINE-94217810; PubMed-8164741;
 RA Bertina R.M., Koelaman B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reijnders P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C."
 RL Nature 369:64-67(1994).
 RN [10]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese."
 RL Blood 91:1135-1139(1998).
 RN [11]
 RP VARIANT APCR THR-334.
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C."
 RL Blood 91:1140-1144(1998).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -1- PTH: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the active
 CC terminus and a light chain at the C-terminus).
 CC -1- PTH: SUFFICIENT IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC -1- DISEASE: OMEN PARAHYEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
 CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGESTS THAT
 CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 CC IMPLANTATION.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L32759; AAB59401.1; JOINED.
 DR EMBL; L32760; AAB59401.1; JOINED.
 DR EMBL; L32761; AAB59401.1; JOINED.
 DR EMBL; L32762; AAB59401.1; JOINED.
 DR EMBL; L32763; AAB59401.1; JOINED.
 DR EMBL; L32764; AAB59401.1; JOINED.
 DR EMBL; L32765; AAB59401.1; JOINED.
 DR EMBL; L32766; AAB59401.1; JOINED.
 DR EMBL; L32767; AAB59401.1; JOINED.
 DR EMBL; L32768; AAB59401.1; JOINED.
 DR EMBL; L32769; AAB59401.1; JOINED.
 DR EMBL; L32770; AAB59401.1; JOINED.
 DR EMBL; L32771; AAB59401.1; JOINED.
 DR EMBL; L32772; AAB59401.1; JOINED.
 DR EMBL; L32773; AAB59401.1; JOINED.
 DR EMBL; L32774; AAB59401.1; JOINED.
 DR EMBL; L32775; AAB59401.1; JOINED.
 DR EMBL; L32776; AAB59401.1; JOINED.
 DR EMBL; L32777; AAB59401.1; JOINED.
 DR EMBL; L32778; AAB59401.1; JOINED.
 DR EMBL; L32779; AAB59401.1; JOINED.
 DR EMBL; L32780; AAB59401.1; JOINED.
 DR EMBL; L32781; AAB59401.1; JOINED.
 DR EMBL; L32782; AAB59401.1; JOINED.
 DR EMBL; L32783; AAB59401.1; JOINED.
 DR EMBL; L32784; AAB59401.1; JOINED.
 DR EMBL; L32785; AAB59401.1; JOINED.
 DR EMBL; L32786; AAB59401.1; JOINED.
 DR EMBL; L32787; AAB59401.1; JOINED.
 DR EMBL; L32788; AAB59401.1; JOINED.
 DR EMBL; L32789; AAB59401.1; JOINED.
 DR EMBL; L32790; AAB59401.1; JOINED.
 DR EMBL; L32791; AAB59401.1; JOINED.
 DR EMBL; L32792; AAB59401.1; JOINED.
 DR EMBL; L32793; AAB59401.1; JOINED.
 DR EMBL; L32794; AAB59401.1; JOINED.
 DR EMBL; L32795; AAB59401.1; JOINED.
 DR EMBL; L32796; AAB59401.1; JOINED.
 DR EMBL; L32797; AAB59401.1; JOINED.
 DR EMBL; L32798; AAB59401.1; JOINED.
 DR EMBL; L32799; AAB59401.1; JOINED.
 DR EMBL; L32800; AAB59401.1; JOINED.
 DR EMBL; L32801; AAB59401.1; JOINED.
 DR EMBL; L32802; AAB59401.1; JOINED.
 DR EMBL; L32803; AAB59401.1; JOINED.
 DR EMBL; L32804; AAB59401.1; JOINED.
 DR EMBL; L32805; AAB59401.1; JOINED.
 DR EMBL; L32806; AAB59401.1; JOINED.
 DR EMBL; L32807; AAB59401.1; JOINED.
 DR EMBL; L32808; AAB59401.1; JOINED.
 DR EMBL; L32809; AAB59401.1; JOINED.
 DR EMBL; L32810; AAB59401.1; JOINED.
 DR EMBL; L32811; AAB59401.1; JOINED.
 DR EMBL; L32812; AAB59401.1; JOINED.
 DR EMBL; L32813; AAB59401.1; JOINED.
 DR EMBL; L32814; AAB59401.1; JOINED.
 DR EMBL; L32815; AAB59401.1; JOINED.
 DR EMBL; L32816; AAB59401.1; JOINED.
 DR EMBL; L32817; AAB59401.1; JOINED.
 DR EMBL; L32818; AAB59401.1; JOINED.
 DR EMBL; L32819; AAB59401.1; JOINED.
 DR EMBL; L32820; AAB59401.1; JOINED.
 DR EMBL; L32821; AAB59401.1; JOINED.
 DR EMBL; L32822; AAB59401.1; JOINED.
 DR EMBL; L32823; AAB59401.1; JOINED.
 DR EMBL; L32824; AAB59401.1; JOINED.
 DR EMBL; L32825; AAB59401.1; JOINED.
 DR EMBL; L32826; AAB59401.1; JOINED.
 DR EMBL; L32827; AAB59401.1; JOINED.
 DR EMBL; L32828; AAB59401.1; JOINED.
 DR EMBL; L32829; AAB59401.1; JOINED.
 DR EMBL; L32830; AAB59401.1; JOINED.
 DR EMBL; L32831; AAB59401.1; JOINED.
 DR EMBL; L32832; AAB59401.1; JOINED.
 DR EMBL; L32833; AAB59401.1; JOINED.
 DR EMBL; L32834; AAB59401.1; JOINED.
 DR EMBL; L32835; AAB59401.1; JOINED.
 DR EMBL; L32836; AAB59401.1; JOINED.
 DR EMBL; L32837; AAB59401.1; JOINED.
 DR EMBL; L32838; AAB59401.1; JOINED.
 DR EMBL; L32839; AAB59401.1; JOINED.
 DR EMBL; L32840; AAB59401.1; JOINED.
 DR EMBL; L32841; AAB59401.1; JOINED.
 DR EMBL; L32842; AAB59401.1; JOINED.
 DR EMBL; L32843; AAB59401.1; JOINED.
 DR EMBL; L32844; AAB59401.1; JOINED.
 DR EMBL; L32845; AAB59401.1; JOINED.
 DR EMBL; L32846; AAB59401.1; JOINED.
 DR EMBL; L32847; AAB59401.1; JOINED.
 DR EMBL; L32848; AAB59401.1; JOINED.
 DR EMBL; L32849; AAB59401.1; JOINED.
 DR EMBL; L32850; AAB59401.1; JOINED.
 DR EMBL; L32851; AAB59401.1; JOINED.
 DR EMBL; L32852; AAB59401.1; JOINED.
 DR EMBL; L32853; AAB59401.1; JOINED.
 DR EMBL; L32854; AAB59401.1; JOINED.
 DR EMBL; L32855; AAB59401.1; JOINED.
 DR EMBL; L32856; AAB59401.1; JOINED.
 DR EMBL; L32857; AAB59401.1; JOINED.
 DR EMBL; L32858; AAB59401.1; JOINED.
 DR EMBL; L32859; AAB59401.1; JOINED.
 DR EMBL; L32860; AAB59401.1; JOINED.
 DR EMBL; L32861; AAB59401.1; JOINED.
 DR EMBL; L32862; AAB59401.1; JOINED.
 DR EMBL; L32863; AAB59401.1; JOINED.
 DR EMBL; L32864; AAB59401.1; JOINED.
 DR EMBL; L32865; AAB59401.1; JOINED.
 DR EMBL; L32866; AAB59401.1; JOINED.
 DR EMBL; L32867; AAB59401.1; JOINED.
 DR EMBL; L32868; AAB59401.1; JOINED.
 DR EMBL; L32869; AAB59401.1; JOINED.
 DR EMBL; L32870; AAB59401.1; JOINED.
 DR EMBL; L32871; AAB59401.1; JOINED.
 DR EMBL; L32872; AAB59401.1; JOINED.
 DR EMBL; L32873; AAB59401.1; JOINED.
 DR EMBL; L32874; AAB59401.1; JOINED.
 DR EMBL; L32875; AAB59401.1; JOINED.
 DR EMBL; L32876; AAB59401.1; JOINED.
 DR EMBL; L32877; AAB59401.1; JOINED.
 DR EMBL; L32878; AAB59401.1; JOINED.
 DR EMBL; L32879; AAB59401.1; JOINED.
 DR EMBL; L32880; AAB59401.1; JOINED.
 DR EMBL; L32881; AAB59401.1; JOINED.
 DR EMBL; L32882; AAB59401.1; JOINED.
 DR EMBL; L32883; AAB59401.1; JOINED.
 DR EMBL; L32884; AAB59401.1; JOINED.
 DR EMBL; L32885; AAB59401.1; JOINED.
 DR EMBL; L32886; AAB59401.1; JOINED.
 DR EMBL; L32887; AAB59401.1; JOINED.
 DR EMBL; L32888; AAB59401.1; JOINED.
 DR EMBL; L32889; AAB59401.1; JOINED.
 DR EMBL; L32890; AAB59401.1; JOINED.
 DR EMBL; L32891; AAB59401.1; JOINED.
 DR EMBL; L32892; AAB59401.1; JOINED.
 DR EMBL; L32893; AAB59401.1; JOINED.
 DR EMBL; L32894; AAB59401.1; JOINED.
 DR EMBL; L32895; AAB59401.1; JOINED.
 DR EMBL; L32896; AAB59401.1; JOINED.
 DR EMBL; L32897; AAB59401.1; JOINED.
 DR EMBL; L32898; AAB59401.1; JOINED.
 DR EMBL; L32899; AAB59401.1; JOINED.
 DR EMBL; L32900; AAB59401.1; JOINED.
 DR EMBL; L32901; AAB59401.1; JOINED.
 DR EMBL; L32902; AAB59401.1; JOINED.
 DR EMBL; L32903; AAB59401.1; JOINED.
 DR EMBL; L32904; AAB59401.1; JOINED.
 DR EMBL; L32905; AAB59401.1; JOINED.
 DR EMBL; L32906; AAB59401.1; JOINED.
 DR EMBL; L32907; AAB59401.1; JOINED.
 DR EMBL; L32908; AAB59401.1; JOINED.
 DR EMBL; L32909; AAB59401.1; JOINED.
 DR EMBL; L32910; AAB59401.1; JOINED.
 DR EMBL; L32911; AAB59401.1; JOINED.
 DR EMBL; L32912; AAB59401.1; JOINED.
 DR EMBL; L32913; AAB59401.1; JOINED.
 DR EMBL; L32914; AAB59401.1; JOINED.
 DR EMBL; L32915; AAB59401.1; JOINED.
 DR EMBL; L32916; AAB59401.1; JOINED.
 DR EMBL; L32917; AAB59401.1; JOINED.
 DR EMBL; L32918; AAB59401.1; JOINED.
 DR EMBL; L32919; AAB59401.1; JOINED.
 DR EMBL; L32920; AAB59401.1; JOINED.
 DR EMBL; L32921; AAB59401.1; JOINED.
 DR EMBL; L32922; AAB59401.1; JOINED.
 DR EMBL; L32923; AAB59401.1; JOINED.
 DR EMBL; L32924; AAB59401.1; JOINED.
 DR EMBL; L32925; AAB59401.1; JOINED.
 DR EMBL; L32926; AAB59401.1; JOINED.
 DR EMBL; L32927; AAB59401.1; JOINED.
 DR EMBL; L32928; AAB59401.1; JOINED.
 DR EMBL; L32929; AAB59401.1; JOINED.
 DR EMBL; L32930; AAB59401.1; JOINED.
 DR EMBL; L32931; AAB59401.1; JOINED.
 DR EMBL; L32932; AAB59401.1; JOINED.
 DR EMBL; L32933; AAB59401.1; JOINED.
 DR EMBL; L32934; AAB59401.1; JOINED.
 DR EMBL; L32935; AAB59401.1; JOINED.
 DR EMBL; L32936; AAB59401.1; JOINED.
 DR EMBL; L32937; AAB59401.1; JOINED.
 DR EMBL; L32938; AAB59401.1; JOINED.
 DR EMBL; L32939; AAB59401.1; JOINED.
 DR EMBL; L32940; AAB59401.1; JOINED.
 DR EMBL; L32941; AAB59401.1; JOINED.
 DR EMBL; L32942; AAB59401.1; JOINED.
 DR EMBL; L32943; AAB59401.1; JOINED.
 DR EMBL; L32944; AAB59401.1; JOINED.
 DR EMBL; L32945; AAB59401.1; JOINED.
 DR EMBL; L32946; AAB59401.1; JOINED.
 DR EMBL; L32947; AAB59401.1; JOINED.
 DR EMBL; L32948; AAB59401.1; JOINED.
 DR EMBL; L32949; AAB59401.1; JOINED.
 DR EMBL; L32950; AAB59401.1; JOINED.
 DR EMBL; L32951; AAB59401.1; JOINED.
 DR EMBL; L32952; AAB59401.1; JOINED.
 DR EMBL; L32953; AAB59401.1; JOINED.
 DR EMBL; L32954; AAB59401.1; JOINED.
 DR EMBL; L32955; AAB59401.1; JOINED.
 DR EMBL; L32956; AAB59401.1; JOINED.
 DR EMBL; L32957; AAB59401.1; JOINED.
 DR EMBL; L32958; AAB59401.1; JOINED.
 DR EMBL; L32959; AAB59401.1; JOINED.
 DR EMBL; L32960; AAB59401.1; JOINED.
 DR EMBL; L32961; AAB59401.1; JOINED.
 DR EMBL; L32962; AAB59401.1; JOINED.
 DR EMBL; L32963; AAB59401.1; JOINED.
 DR EMBL; L32964; AAB59401.1; JOINED.
 DR EMBL; L32965; AAB59401.1; JOINED.
 DR EMBL; L32966; AAB59401.1; JOINED.
 DR EMBL; L32967; AAB59401.1; JOINED.
 DR EMBL; L32968; AAB59401.1; JOINED.
 DR EMBL; L32969; AAB59401.1; JOINED.
 DR EMBL; L32970; AAB59401.1; JOINED.
 DR EMBL; L32971; AAB59401.1; JOINED.
 DR EMBL; L32972; AAB59401.1; JOINED.
 DR EMBL; L32973; AAB59401.1; JOINED.
 DR EMBL; L32974; AAB59401.1; JOINED.
 DR EMBL; L32975; AAB59401.1; JOINED.
 DR EMBL; L32976; AAB59401.1; JOINED.
 DR EMBL; L32977; AAB59401.1; JOINED.
 DR EMBL; L32978; AAB59401.1; JOINED.
 DR EMBL; L32979; AAB59401.1; JOINED.
 DR EMBL; L32980; AAB59401.1; JOINED.
 DR EMBL; L32981; AAB59401.1; JOINED.
 DR EMBL; L32982; AAB59401.1; JOINED.
 DR EMBL; L32983; AAB59401.1; JOINED.
 DR EMBL; L32984; AAB59401.1; JOINED.
 DR EMBL; L32985; AAB59401.1; JOINED.
 DR EMBL; L32986; AAB59401.1; JOINED.
 DR EMBL; L32987; AAB59401.1; JOINED.
 DR EMBL; L32988; AAB59401.1; JOINED.
 DR EMBL; L32989; AAB59401.1; JOINED.
 DR EMBL; L32990; AAB59401.1; JOINED.
 DR EMBL; L32991; AAB59401.1; JOINED.
 DR EMBL; L32992; AAB59401.1; JOINED.
 DR EMBL; L32993; AAB59401.1; JOINED.
 DR EMBL; L32994; AAB59401.1; JOINED.
 DR EMBL; L32995; AAB59401.1; JOINED.
 DR EMBL; L32996; AAB59401.1; JOINED.
 DR EMBL; L32997; AAB59401.1; JOINED.
 DR EMBL; L32998; AAB59401.1; JOINED.
 DR EMBL; L32999; AAB59401.1; JOINED.
 DR EMBL; L33000; AAB59401.1; JOINED.
 DR EMBL; L33001; AAB59401.1; JOINED.
 DR EMBL; L33002; AAB59401.1; JOINED.
 DR EMBL; L33003; AAB59401.1; JOINED.
 DR EMBL; L33004; AAB59401.1; JOINED.
 DR EMBL; L33005; AAB59401.1; JOINED.
 DR EMBL; L33006; AAB59401.1; JOINED.
 DR EMBL; L33007; AAB59401.1; JOINED.
 DR EMBL; L33008; AAB59401.1; JOINED.
 DR EMBL; L33009; AAB59401.1; JOINED.
 DR EMBL; L33010; AAB59401.1; JOINED.
 DR EMBL; L33011; AAB59401.1; JOINED.
 DR EMBL; L33012; AAB59401.1; JOINED.
 DR EMBL; L33013; AAB59401.1; JOINED.
 DR EMBL; L33014; AAB59401.1; JOINED.
 DR EMBL; L33015; AAB59401.1; JOINED.
 DR EMBL; L33016; AAB59401.1; JOINED.
 DR EMBL; L33017; AAB59401.1; JOINED.
 DR EMBL; L33018; AAB59401.1; JOINED.
 DR EMBL; L33019; AAB59401.1; JOINED.
 DR EMBL; L33020; AAB59401.1; JOINED.
 DR EMBL; L33021; AAB59401.1; JOINED.
 DR EMBL; L33022; AAB59401.1; JOINED.
 DR EMBL; L33023; AAB59401.1; JOINED.
 DR EMBL; L33024; AAB59401.1; JOINED.
 DR EMBL; L33025; AAB59401.1; JOINED.
 DR EMBL; L33026; AAB59401.1; JOINED.
 DR EMBL; L33027; AAB59401.1; JOINED.
 DR EMBL; L33028; AAB59401.1; JOINED.
 DR EMBL; L33029; AAB59401.1; JOINED.
 DR EMBL; L33030; AAB59401.1; JOINED.
 DR EMBL; L33031; AAB59401.1; JOINED.
 DR EMBL; L33032; AAB59401.1; JOINED.
 DR EMBL; L33033; AAB59401.1; JOINED.
 DR EMBL; L33034; AAB59401.1; JOINED.
 DR EMBL; L33035; AAB59401.1; JOINED.
 DR EMBL; L33036; AAB59401.1; JOINED.
 DR EMBL; L33037; AAB59401.1; JOINED.
 DR EMBL; L33038; AAB59401.1; JOINED.
 DR EMBL; L33039; AAB59401.1; JOINED.
 DR EMBL; L33040; AAB59401.1; JOINED.
 DR EMBL; L33041; AAB59401.1; JOINED.
 DR EMBL; L33042; AAB59401.1; JOINED.
 DR EMBL; L33043; AAB59401.1; JOINED.
 DR EMBL; L33044; AAB59401.1; JOINED.
 DR EMBL; L33045; AAB59401.1; JOINED.
 DR EMBL; L33046; AAB59401.1; JOINED.
 DR EMBL; L33047; AAB59401.1; JOINED.
 DR EMBL; L33048; AAB59401.1; JOINED.
 DR EMBL; L33049; AAB59401.1; JOINED.
 DR EMBL; L33050; AAB59401.1; JOINED.
 DR EMBL; L33051; AAB59401.1; JOINED.
 DR EMBL; L33052; AAB59401.1; JOINED.
 DR EMBL; L33053; AAB59401.1; JOINED.
 DR EMBL; L33054; AAB59401.1; JOINED.
 DR EMBL; L33055; AAB59401.1; JOINED.
 DR EMBL; L33056; AAB59401.1; JOINED.
 DR EMBL; L33057; AAB59401.1; JOINED.
 DR EMBL; L33058; AAB59401.1; JOINED.
 DR EMBL; L33059; AAB59401.1; JOINED.
 DR EMBL; L33060; AAB59401.1; JOINED.
 DR EMBL; L33061; AAB59401.1; JOINED.
 DR EMBL; L33062; AAB59401.1; JOINED.
 DR EMBL; L33063; AAB59401.1; JOINED.
 DR EMBL; L33064; AAB59401.1; JOINED.
 DR EMBL; L33065; AAB59401.1; JOINED.
 DR EMBL; L33066; AAB59401.1; JOINED.
 DR EMBL; L33067; AAB59401.1; JOINED.
 DR EMBL; L33068; AAB59401.1; JOINED.
 DR EMBL; L33069; AAB59401.1; JOINED.
 DR EMBL; L33070; AAB59401.1; JOINED.
 DR EMBL; L33071; AAB59401.1; JOINED.
 DR EMBL; L33072; AAB59401.1; JOINED.
 DR EMBL; L33073; AAB59401.1; JOINED.
 DR EMBL; L33074; AAB59401.1; JOINED.
 DR EMBL; L33075; AAB59401.1; JOINED.
 DR EMBL; L33076; AAB59401.1; JOINED.
 DR EMBL; L33077; AAB59401.1; JOINED.
 DR EMBL; L33078; AAB59401.1; JOINED.
 DR EMBL; L33079; AAB59401.1; JOINED.
 DR EMBL; L33080; AAB59401.1; JOINED.
 DR EMBL; L33081; AAB59401.1; JOINED.
 DR EMBL; L33082; AAB59401.1; JOINED.
 DR EMBL; L33083; AAB59401.1; JOINED.
 DR EMBL; L33084; AAB59401.1; JOINED.
 DR EMBL; L33085; AAB59401.1; JOINED.
 DR EMBL; L33086; AAB59401.1; JOINED.
 DR EMBL; L33087; AAB59401.1; JOINED.
 DR EMBL; L33088; AAB59401.1; JOINED.
 DR EMBL; L33089; AAB59401.1; JOINED.
 DR EMBL; L33090; AAB59401.1; JOINED.
 DR EMBL; L33091; AAB59401.1; JOINED.
 DR EMBL; L33092; AAB59401.1; JOINED.
 DR EMBL; L33093; AAB59401.1; JOINED.
 DR EMBL; L33094; AAB59401.1; JOINED.
 DR EMBL; L33095; AAB59401.1; JOINED.
 DR EMBL; L33096; AAB59401.1; JOINED.
 DR EMBL; L33097; AAB59401.1; JOINED.
 DR EMBL; L33098; AAB59401.1; JOINED.
 DR EMBL; L33099; AAB59401.1; JOINED.
 DR EMBL; L33100; AAB59401.1; JOINED.
 DR EMBL; L33101; AAB59401.1; JOINED.
 DR EMBL; L33102; AAB59401.1; JOINED.
 DR EMBL; L33103; AAB59401.1; JOINED.
 DR EMBL; L33104; AAB59401.1; JOINED.
 DR EMBL; L33105; AAB59401.1; JOINED.
 DR EMBL; L33106; AAB59401.1; JOINED.
 DR EMBL; L33107; AAB59401.1; JOINED.
 DR EMBL; L33108; AAB59401.1; JOINED.
 DR EMBL; L33109; AAB59401.1; JOINED.
 DR EMBL; L33110; AAB59401.1; JOINED.
 DR EMBL; L33111; AAB59401.1; JOINED.
 DR EMBL; L33112; AAB59401.1; JOINED.
 DR EMBL; L33113; AAB59401.1; JOINED.
 DR EMBL; L33114; AAB59401.1; JOINED.
 DR EMBL; L33115; AAB59401.1; JOINED.
 DR EMBL; L33116; AAB59401.1; JOINED.
 DR EMBL; L33117; AAB59401.1; JOINED.
 DR EMBL; L33118; AAB59401.1; JOINED.
 DR EMBL; L33119; AAB59401.1; JOINED.
 DR EMBL; L33120; AAB59401.1; JOINED.
 DR EMBL; L33121; AAB59401.1; JOINED.
 DR EMBL; L33122; AAB59401.1; JOINED.
 DR EMBL; L33123; AAB59401.1; JOINED.
 DR EMBL; L33124; AAB59401.1; JOINED.
 DR EMBL; L33125; AAB59401.1; JOINED.
 DR EMBL; L33126; AAB59401.1; JOINED.
 DR EMBL; L33127; AAB59401.1; JOINED.
 DR EMBL; L33128; AAB59401.1; JOINED.
 DR EMBL; L33129; AAB59401.1; JOINED.
 DR EMBL; L33130; AAB59401.1; JOINED.
 DR EMBL; L33131; AAB

```

FT REPEAT 1302 1310 2-14.

Query Match 8.4%; Score 317; DB 1; Length 2224;
Best Local Similarity 42.0%; Pred. No. 4.4e-14;
Matches 71; Conservative 30; Misses 52; Indels 16; Gaps 5;

OY 253 NCSCSLSPF-----PDGQIRASS---SMQSVNMSGQVHMSGGARLDQGPSSAGDSS 305
    ||||| : ||||| || | ||||| || ||||| || ||||| || |||||
DB 2064 NCCSTPLGHEKNGKIENKOITASSFFKSW----GD--TWEPFRALNAQGVNMAQA 2116
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 306 NNHKREWLEIDLGKKRTIGTGTSGSNFNPFYKSFVNFKNNNSKMKRYKGIVNNE 365
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2117 NNNK--QMTEIDLTKIKRTAITIITQGCKSLSEMYVKSTIHHSQGVEMKRKYRLKSSAV 2174
    ::||::| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 366 EAVFGQNSNFDPVONNPPIPVARYRVVPOTWHORIALKVELLGCQT 414
    ::||::| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2175 DKIEFGNTWKGVNKFNFNPITSFRIVRIPKTWNOSITLRLEFGCDI 2223
    ::||::| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
FA5_PIG STANDARD: PRT, 2258 AA.
AC Q9GLP1;
DT 16-OCT-2001 (Rel. 40, created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9623;
[1]
SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RA MEDLINE-21121490; PubMed-1129814;
RX Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.,
"Porcine factor V: cDNA cloning, gene mapping, three-dimensional
protein modeling of membrane binding sites and comparative anatomy of
domains."
RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -1- FUNCTION: Coagulation factor V is a cofactor that participates
with factor Xa to activate prothrombin to thrombin.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
chain, noncovalently bound. The interaction between the two chains
is calcium-dependent.
CC -1- DOMAIN: Domain B contains 41 x 9 AA tandem repeats. Domains C1
and C2 may be involved in membrane binding.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
cofactor, factor Va (formation of a heavy chain at the N-terminus and a light chain at the C-terminus).
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AF191308; AAC28381.1; -.
DR HSSP; P12259; ICZT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.

```

[illegible]

[illegible]

	Query Match	8.3% ; Score 314; DB 1;
	Best Local Similarity	39.2%; Pred. No. 7.3e-14;
	Matches 67; Conservative	35; Mismatches 49; Indels 20; Gaps 6;
QY	253 NCGRSLSTFE-----PDGQIRASS--SWQSVNESCIDQVHMSPGARLDDQG--PSMASGD 303	: : : : : : :
Dc	2098 NGCSTPLMGESGNINKEDITASSFRKSMW-----GD--TWEPFRRRLNAOGHVANAMOKA 2150	: :
QY	304 SSNNHKPREMLEIDLGEEKKITGTIRTGSTOSNFNVKSFPMNKNNNSKKVTYKGIVN 363	::: : : : : : : : : : : : : : : : :
Dc	2151 NNNN---QWQIDLTKRKKTITAITTGCKSKLSISMYYKRRTIQSDSGVEKSRREKSS 2206	: :
QY	364 NEERVFQGSNFRDPVONNPPIPVARYRVVPQTMRORLAKVELIGCOI 414	: : : : : : : : : : : : : : : :
Dc	2207 MYDKTFEGNNNIKGVKNFFNPPIISRIRILPKMMNQSIHLRELFGCDI 2257	: : : : : : : : : : : : : : : :
RESULT 12		
MFCM_HUMAN		
ID	MFCM_HUMAN	STANDARD; PRT; 387 AA.
AC	008431:	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Lactadherin precursor (Milk fat globule-ECF factor 8) (MFG-E8) (HMFG)	
DE	(breast epithelial antigen BA46) (MFCM) [Contains: Medin].	
GN	MFC8.	
OS	Homo sapiens (Human).	

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Certani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urra R., Kunlyoshi J., Bistrain A.M.,
RA Certani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Guiffrida M.G., Cavalletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RC TISSUE=Milk;
RX MEDLINE=99342076; PubMed=10411933;
RA Haegggvist B., Naesslund J., Stetten K., Westermarck G.T., Mucchiano G.,
RA Thernberg L.O., Nordstam C., Engstrom U., Westermarck P.;
RT "Medin, an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Certani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";
RL DNA Cell Biol. 16:861-869(1997).
CC -1- FUNCTION. MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC -1- FUNCTION. MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC -1- SUBCELLULAR LOCATION. PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY. MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC -1- PGM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC AMINO ACID 264 AND 273.
CC -1- SIMILARITY. CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY. CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license@lsb.slb.ch).
CC or send an email to license@lsb.slb.ch).
CC -----
DR EMBL: U58516; AAC50549.1; -
DR EMBL: S56151; AB19771.1; -
DR HSPR: P08709; IBF9.
DR Genew: HGNC:7036; MFGEB.
DR MIM: 602281; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000421; FAS5_C.
DR Pfam: PFO00008; EGF: 1.
DR

RESULT 14
 ID FA8_HUMAN STANDARD; PRT; 2351 AA.
 AC P00451;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor VIII precursor (Procoagulant component)
 DE (Antihemophilic factor) (AHF).
 GN F8 OR F8C.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86081164; PubMed=3935400;
 RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
 RA Hartog K., Kuo C.H., Maslarsz F.R., Merryweather J.P., Najarian R.,
 RA Pechl C., Potter S.J., Puma J., Outroga M., Rall L.B., Randolph A.,
 RA Urdas M.S., Valenzuela P., Dahl H.-H.M., Favaloro J., Hansen J.,
 RA Nordfang O., Ezban M.
 RT Characterization of the polypeptide composition of human factor
 RT VIII:C and the nucleotide sequence and expression of the human kidney
 RT cDNA.";
 RL DNA 4:333-349(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061548; PubMed=6438526;
 RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschler J.,
 RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion R.L.,
 RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
 RT Expression of active human factor VIII from recombinant DNA clones.";
 RL Nature 312:330-337(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061550; PubMed=6438528;
 RA Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
 RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
 RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,
 RA Hewick R.M.;
 RT Molecular cloning of a cDNA encoding human antihemophilic factor.";
 RL Nature 312:342-347(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93265012; PubMed=1303178;
 RA Gitschler J., Wood W.I.;
 RT "Sequence of the exon-containing regions of the human factor VIII
 RT gene.";
 RL Hum. Mol. Genet. 1:199-200(1992).
 RN [5]
 RP SEQUENCE OF 2064-2070 FROM N.A.
 RA de Water N.S., Williams R., Browett P.J.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SULFATION OF TYR-1699.
 RX MEDLINE=91093266; PubMed=1898735;
 RA Leyte A., van Schijndel H.B., Niehs C., Hutner W.B., Verbeet M.P.,
 RA Merens K., van Mourik J.A.;
 RT Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RP SULFATION.
 RX MEDLINE=92207952; PubMed=1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3325(1992).
 RN [8]

RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE=95200924; PubMed=7893714;
 RA Gilbert G.E., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=91221499; PubMed=1902642;
 RA Gitschler J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=89088506; PubMed=2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=95245332; PubMed=7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT GLN-2326.
 RX MEDLINE=86235434; PubMed=3012775;
 RA Gitschler J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic.";
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT PRO-2135.
 RX MEDLINE=88096539; PubMed=3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschler J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT GLN-2228.
 RX MEDLINE=88191889; PubMed=2833855;
 RA Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE=88220354; PubMed=2835904;
 RA Yousoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE=89274393; PubMed=2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE=90001543; PubMed=2506948;
 RA Shima M., Ware J., Yoshioke A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE=90057860; PubMed=2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;

FT	DOMAIN	308	463	F5/8 TYPE C 2.
FT	SITE	87	89	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	28	39	BY SIMILARITY.
FT	DISULFID	33	49	BY SIMILARITY.
FT	DISULFID	51	60	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	96	BY SIMILARITY.
FT	DISULFID	98	107	BY SIMILARITY.
FT	DISULFID	148	303	BY SIMILARITY.
FT	DISULFID	290	294	BY SIMILARITY.
FT	CARBOHYD	308	463	BY SIMILARITY.
FT	CARBOHYD	61	61	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	266	266	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	426	426	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	30	30	S -> F (IN REF. 2).
FT	CONFLICT	35	35	N -> D (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	110	147	ETRYNLDGEHMTTAVPNTAVPTPAPDLSNNLSR ->
FT	CONFLICT	168	168	G (IN REF. 2).
FT	CONFLICT	196	196	Y -> S (IN REF. 2).
FT	CONFLICT	309	309	H -> T (IN REF. 2).
FT	CONFLICT	395	395	L -> S (IN REF. 2).
FT	CONFLICT	395	395	E -> A (IN REF. 2).
SO	SEQUENCE	463 AA;	51465 MW;	D78B6CEFBBA724D CRC64;

Query Match 7.38; Score 276.5; DB 1; Length 463;
 Best Local Similarity 23.48; Pred. No. 3.4e-12;
 Matches 112; Conservative 59; Mismatches 118; Indels 189; Gaps 21;

QY	94	CASDYLLFTSSDOYGRYSSMTVPKELLANTSEVYVRFSSGSHISR-----GF 143
DB	16	CASG--LEAASGD-----FCDS-----SLCLN-----GGTCLTGQDNDIYCLCEPGEF 55
QY	144	LTLYAS-----SDHDLITCLERASHYLKTEYSKFCPCAGCRVAGDISGM- 189
DB	56	TGLVCNETERPCSPNCPYDARCLVT-IDTQNGDITETICQCPVGSIGHCETETNY 114
QY	190	-VDG-YRDTSLCKAAI-----HAGITAD----- 211
DB	115	NLDGEYMTTAVPNTAVPTPAPDLSNNLSRCSPTQLGMEGAIADSOISASYVMGFM 174
QY	212	--ELGGQISVLRKGI-----SRYEGILANGVLSRDG----- 241
DB	175	GLQRMGPDLRLRYRTGIVNANMAHSNYSKFWIQLRKMRVSGVMTQGA-SRAGRAEYL 233
QY	242	-----SLSDKRRLFTSN-----GCSR-- 257
DB	234	KTFKVAISLDGRKFEFIQDESQDKEFLGNDLNNLSLKVNFPTLEAOYIRLTPVSGCHRG 293
QY	258	-SLSFE-----PDGQIRASSSMQSVNESGDQVHWSFGQARLDQDQ 296
DB	294	CTLRFELLCGCLHGLEPLGLKNNITIPDSQMSASSSYKWTNLR--FGWYPHLGRLDNQG 351
QY	297	P--SNASGDSNNHKKPREWEIDLEKKTITRTGSGTOSNPNFYKSVFVNFKNNSK 354
DB	352	KINAM-----TAQSNASAKEMLDGTFQROVTLITOGARDFGHIOYVESYKVAHSDGVO 407
QY	355	WKTYKGIIVNNEKRVFQNSNFRDPVONNFIPIVAVRVVVPOTWHORIALKVELIGC 412
DB	408	WTYE--EGSSSKVFGQNDLNNHKKNIPEKPMARVAVRLPVSWHNRITLRLELLGC 463

Search completed: May 15, 2003, 13:16:06
 Job time : 32.3719 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 69,8563 Seconds
(without alignments)
2108.953 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGARGGALARAAGRL.....YSAPROCLPLNQTAFALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-mhc:*
- 8: SP-organellar:*
- 9: SP-phage:*
- 10: SP-plant:*
- 11: SP-rodent:*
- 12: SP-virus:*
- 13: SP-vertebrate:*
- 14: SP-unclassified:*
- 15: SP-virus:*
- 16: SP-bacteriophage:*
- 17: SP-archaeo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1925.5	51.1	503	11 09DAJ3	09d4j3 mus musculus
2	1769.5	46.9	460	11 09D696	09d696 mus musculus
3	1535	40.7	432	11 08R327	08r327 mus musculus
4	1082	28.2	769	11 091ZV3	091zv3 mus musculus
5	1055	28.0	769	11 091ZV2	091zv2 mus musculus
6	1042	27.6	775	4 096PD2	096pd2 homo sapien
7	1042	27.6	775	4 096PD2	096pd2 homo sapien
8	841	22.3	155	4 096NM2	096nm2 homo sapien
9	572.5	15.2	364	4 014089	014089 homo sapien
10	357.5	9.3	926	11 08QZ17	08qz17 mus musculus
11	351.5	9.3	901	4 09H2E4	09h2e4 homo sapien
12	351.5	9.3	901	4 09H2D5	09h2d5 homo sapien
13	351.5	9.3	906	4 09H2E3	09h2e3 homo sapien
14	351.5	9.3	906	4 09H2D4	09h2d4 homo sapien
15	351	9.3	919	13 08UVR0	08uvr0 gallus galli
16	351	9.3	936	13 08UVR0	08uvr0 gallus galli

17	349	9.3	555	4 09H2E2	09h2e2 homo sapien
18	335	8.9	609	4 096190	096190 homo sapien
19	335	8.9	644	4 0961H5	0961h5 homo sapien
20	335	8.9	704	4 09H2E1	09h2e1 homo sapien
21	325.5	8.6	921	11 09QX38	09qx38 ratu norv
22	321	8.5	2224	4 043737	043737 homo sapien
23	320	8.5	923	13 08QFX6	08qfx6 brachydanto
24	318	8.4	2183	11 088783	088783 mus musculus
25	306.5	8.1	2343	6 018806	018806 carls faml
26	305.5	8.1	2343	6 062730	062730 carls faml
27	299.5	7.9	779	4 09NTT3	09ntt3 homo sapien
28	290	7.7	426	11 09WTS3	09wts3 mus musculus
29	285.5	7.6	216	4 014286	014286 homo sapien
30	279.5	7.4	463	11 09R1X9	09r1x9 mus musculus
31	277	7.3	2119	13 090X47	090x47 brachydanto
32	268	7.1	480	4 043854	043854 homo sapien
33	264	7.0	480	11 035474	035474 mus musculus
34	238.5	6.3	363	6 077718	077718 equus cabal
35	226	6.0	1128	11 088442	088442 mus musculus
36	221.5	5.9	224	11 09R1M6	09r1m6 mus musculus
37	217	5.8	1158	4 014113	014113 homo sapien
38	216.5	5.7	719	11 061281	061281 mus musculus
39	209.5	5.6	728	6 097567	097567 bos taurus
40	200	5.3	1311	4 08WX98	08wx98 homo sapien
41	199.5	5.3	858	5 076470	076470 lytechinus
42	199	5.3	721	4 096M80	096m80 homo sapien
43	194.5	5.2	335	4 09BTL9	09btl9 homo sapien
44	188.5	5.0	745	4 096M75	096m75 homo sapien
45	188.5	5.0	3843	5 09U5D0	09u5d0 drosophila

ALIGNMENTS

RESULT 1	ID	SEQUENCE FROM N.A.	PRELIMINARY	PRT	503 AA.
AC	09DAJ3	09DAJ3			
DC	01-JUN-2001 (TREMBLrel. 17, Created)				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	4631413K1R1K protein.				
GN	4631413K1R1K.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shidate K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,				
RA	Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadoya K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Glass C., King B., Koehli H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schimi L. M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,				
RA	Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D. A., Kamuya M., Lee N. H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shidate Y., Storch K.-F.,				
RA	Suzuki H., Toyokawa K., Wang K. H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,				
RA	Hashizaki Y.;				
RT	*Functional annotation of a full-length mouse cDNA collection.*;				
RL	Nature 409:685-690(2001).				
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.				
DR	EMBL; AK016485; BAB30265.1; -.				

MD: MGI:1913936; 4631413K1IRK.
 DR InterPro: IPR000859; COB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; COB; 1.
 DR SMART: SM00042; COB; 1.
 DR PROSITE: PS01180; COB; 1.
 SO SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;

Query Match 51.18; Score 1925.5; DB 11; Length 503;
 Best Local Similarity 55.88; Pred. No. 6.9e-140;
 Matches 392; Conservative 34; Mismatches 73; Indels 203; Gaps 4;

14 AAGGGLALLAVANAPLRLQAELEGDGGHLYVYDSSSTMTSKNPGYVPHVTECEKIT 73
 5 AGGVSVALLEAVCAPRLQAEELEGDGGHLYVYDSSSTMTSKNPGYVPHVTECEKIT 64
 74 VPKGRRLRLRGLDIEQSCADYLITFTSSSDQPGYCSMTVPKELLMTSEVTFE 133
 65 VPKGRRLRLRGLDIEQSCADYLITFTSSSDQPGYCSMTVPKELLMTSEVTFE 124
 134 SGSHISGRGFLITYASSDHPDLITCLERASHYLTETSKFCPCAGCDVAGDISGMVVDGY 193
 125 SGSHISGRGFLITYASSDHPDLITCLERASHYLTETSKFCPCAGCDVAGDISGMVVDGY 184
 194 RDTSLICAAHAGITDELGGQISYLRKGISREGILANGVLSRGSISDREPLTSN 253
 185 RDTSLICAAHAGITDELGGHINLQSKGISHTGELANGVLSRHSISLEKRLP 241
 254 GCSRSLSFEPDQIRASSSMQSVESDQVHMSPOQARIADQGSWASGDSNNHKKPREW 313
 242 241
 314 LEIDGEEKKITGTGRTTSGTOSNFYVSKVFNKNNKMKYKGIYNNEKVPQNS 373
 242 241
 374 NFRDPVQNNFPIVARYRVVPTWHQRIALKEVLLCQITQGDLSLWRTKSSTSVS 433
 242 241
 434 TKREDETIRPSEBETSTGINTTVAIPLVLLVVFAGMIFPAFAFKKKKSPYSA 493
 242 241
 494 EAQTDCKQIKYPPARHQSAEFTISYDNEKEMTKDILITSDMADYQOPLMIGTVTR 553
 286 DAQTSCKQIKYPPARHQSAEFTISYDNEKEMTKDILITSDMADYQOPLMIGTVTR 345
 554 KGSFRRPMDTAEAGYSTDAGHYDCQORAGRHEHYALPLAPREVTPTVERHVLRAH 613
 346 KGSFRRPMDTAEAGYSTDAGHYDCQORAGRHEHYALPLAPREVTPTVERHVLRAH 405
 614 TFSAQSGRVVPQPGHSHLSGSGFSFVAGVGAQDQYQPHSAQPADRGYDPRKAVSA 673
 406 TFSQSGRVVPQPGHSHLSGSGFSFVAGVGAQDQYQPHSAQPADRGYDPRKAVSA 464
 674 LATESGHPDSQRPPTNPGTSDYSAPRDCPLPLQOTATALL 715
 465 L-DSDRPASOSQMT-SGGDGYGAPRNGLAPLQNTATALL 503

RESULT 2

Q9D696 PRELIMINARY; PRT; 460 AA.
 AC Q9D696;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE 4631413K1IRK.
 GN 4631413K1IRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ichii Y.,
 Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 Guenclinet S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohatsu S.,
 Hayshtizaki Y.;
 FT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 COB DOMAIN.
 DR EMBL: AK014521; BAB29409.1; -
 DR MGD: MGI:1913936; 4631413K1IRK.
 DR InterPro: IPR000859; COB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; COB; 1.
 DR SMART: SM00042; COB; 1.
 DR PROSITE: PS01180; COB; 1.
 SO SEQUENCE 460 AA; 50334 MW; 3CF3356F70A8B993 CRC64;

Query Match 46.98; Score 1769.5; DB 11; Length 460;
 Best Local Similarity 54.48; Pred. No. 6.4e-128;
 Matches 361; Conservative 32; Mismatches 67; Indels 203; Gaps 4;

53 MTSKNYPTGYNHTVCEKITVPPKGRRLRLRGLDIEQSCADYLITFTSSDQPGY 112
 1 MTSKNYPTGYNHTVCEKITVPPKGRRLRLRGLDIEQSCADYLITFTSSDQPGY 60
 113 GSNVPEKELLNTESEVTVREFSGSHISGRGFLITYASSDHPDLITCLERASHYLTETSK 172
 61 GSNVPEKELLNTESEVTVREFSGSHISGRGFLITYASSDHPDLITCLERASHYLTETSK 120
 173 FCPAGCRDVAGDISGMVVDGYQDRTSLICAAHAGITDELGGQISYLRKGISREGIL 232
 121 FCPAGCRDVAGDISGMVVDGYQDRTSLICAAHAGITDELGGHINLQSKGISHTGEL 180
 233 ANGVLSRDSLSDKRFLFTSNGCSRSLSFEPDQIRASSSMQSVESDQVHMSPOQARL 292
 181 ANGVLSRDSLSDKRFLFTSNGCSRSLSFEPDQIRASSSMQSVESDQVHMSPOQARL 198
 293 ODQGSMAQSDSNNHKKPREWLEIDGEEKKITGTGRTTSGTOSNFYVSKVFNKNNN 352
 199 198
 353 SKWKTYGIVNNEKVPQNSNFRDPVQNNFPIVARYRVVPTWHQRIALKEVLLCQIT 412
 199 198
 413 QITQGDLSLWRTKSSTSVTKREDETIRPSEBETSTGINTTVAIPLVLLVVF 472
 199 198
 473 GNGIFAAFRKKKKKSPYSAEAOQTDCKQIKYPPARHQSAEFTISYDNEKEMTKDIL 532
 222 GNGIFAAFRKKKKKSPYSAEAOQTDCKQIKYPPARHQSAEFTISYDNEKEMTKDIL 281
 533 ITSDMADYQOPLMIGTVTRKGSFRRPMDTAEAGYSTDAGHYDCQORAGRHEHYALP 592

Dd	282	ITSDMADYQOPLMIGTGVARKGSTFRPMDTDEEVRAVNTSEASHGYDCPHRGREHYALP	341
Oy	553	LAPPEPVATPTVEKRVLRATHFFSAQSQRVGPQGKHSHSSGGFSFVAVGADGDY	652
Dd	342	LTHSPEVATPTVEKRVLRATHFFSQSGRYVGPGRTHSHSSGGFPFATG-ATQVESY	400
Oy	653	ORPHSAQPADRCYDRBRKAVSATATESGHHDSQKPTHPCTSDSYGAPRDLPLNOTANT	712
Dd	401	ORPAPRKPRVGGYDAPPAASSFL--DSRPDAAGSQMT-SGDDGYSAPRRGLAFLNQTAHT	457
Oy	713	ALL 715	
Dd	458	ALL 460	
 RESULT 3			
ID	08R327	PRELIMINARY;	PRT; 432 AA.
AC	08R327;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)		
DE	Similar to RIKEN CDNA 4631413K11 gene.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Straubenberg R.;		
RL	Submitted (Apr-2002) to the EMBL/GenBank/DDBJ databases.		
SR	EMBL; BC026771.1; AAH26771.1; -		
DQ	SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD917 CRC64;		
 Query Match 40.7%; Score 1535; DB 11; Length 432; Best Local Similarity 47.0%; Pred. No. 6.9e-110; Matches 330; Conservative 26; Mismatches 72; Indels 274; Gaps			
Oy	14	AAGRGLALLLVASAPRLQAEELDGCGHLVTYQDSGTMTSKNPPTYPNHTVCERTIT	73
Dd	5	AGGPEVLALLFAVCAPRLQAEEI-----	28
Oy	74	VPKGRLLRLRLGDIDIEQTCAVDLLFTSSSDQGYPGYCSMTVPKELLNTEVTVPRE	133
Dd	29	-----GPGYCSMAVPERKELNLNSNEVTVLEK	53
Oy	134	SGSHISGRGFLLTYVASSDPDLITCLERASHLYLKTEYSKFCACGRDVAGDISGMNVDG	193
Dd	54	SGSHISGRGFLLTYVASSDPDLITCLEHGSHFEKEYSKFCACGRDLAGDISGMTKCY	113
Oy	194	RDTSLCAAIAHAGITADELGQISVYLORKGISRYEGILANGCVLSRDGSLDKREFLTSM	253
Dd	114	RDTSLCAAIAHAGITADELGCHIMLQSKGISHEYGELANGCVLSRHGSLSEKREFL---	170
Oy	254	GCSRSLSPEDQIIASSWSQSVNSGDQVHMSPGARLQDQGPSMASGDSSNNHKPREM	313
Dd	171	-----	170
Oy	314	LEIDGERKKKITGITTSQSTOSNFNFYVKSFYMFNRKNNSKWKTKGIYNEEKVFOGNS	373
Dd	171	-----	170
Oy	374	NFRDVOQNPFIPYARAYRVPPQWHRIALAKVELICQLITQGDLSLVKRTSOSTSYS	433
Dd	171	-----	170
Oy	434	TKEDETITRPIPETSETSTGINITVAIPVLVVLFVFGMGIFAFAFKKKKKSGPYGSA	493
Dd	171	-----TPGNMITTVAIPVSIFIALILLTGMSIFAIACRRKKKKGNPYSA	214
Oy	494	EAQTFDCMKQIQYPPARHOSAFTTISYDNKKEMTKLILITSMDADYOOLMIGTGYTR	553
Dd	215	DAQTKCKMQIQYPPARHOSSTETITSYDNKKEMTKLILITSMDADYOOLMIGTGYTR	274

Oy	554	KGSTRPMDOAEAGVSTADGACGHDCPORAGRHEVALPLAPEPEYAPPIVERHYLRAM	613
Dd	275	KGSTRRPMDTDIEEVKRVNTEASGHNDCCPRRGRHETALPLTISEEYPATPIYERHLRLRH	334
Oy	614	TFSAGSYRVGPORPOEHKHSLSSGGFSPVAVGAODGYDRPHSAOPADRGRDRAVA	673
Dd	335	TFSTOSGVPPRPRTLHKSHSGGFPPATG-ATQIESYQRASPAPVPVGXDKPAAASF	393
Oy	674	LATESGHPDSOKPRPHPTSDYSAPROCLTLINOTAMALL	715
Dd	394	I--DSRDPA SOSQMT-SGDDGSAPRCGLAPLOTAMTALL	432

RESULT 4

ID	091ZV3	PRELIMINARY;	PRT:	769 AA.
AC	091ZV3:			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)			
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.			
GN	ESDN.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCHI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
SC	STRAIN=ICR;			
RX	PubMed-11447234;			
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.; Cloned from Vascular ES-DN, a Novel Neuropilin-Like Membrane Protein Expressed by Endothelial Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is Up-regulated after Vascular Injury."			
RT	J. Biol. Chem. 276:34105-34114(2001).			
RL	-1 SIMILARITY: BELONGS TO THE CRP/FMR FAMILY OF TRANSCRIPTIONAL REGULATORS.			
CC	EMBL; AF87548; AACJ0179.1; --			
DR	MGD; MG1:1920629; Esdn.			
DR	InterPro: IPR000859; CUB_domain.			
DR	InterPro: IPR000421; FA58_C.			
DR	InterPro: IPR004043; LCCL_dom.			
DR	Pfam; PF00431; CUB_1.			
DR	Pfam; PF00754; F5_F8_type-C; 1.			
DR	PROSITE; PS01180; CUB; 1.			
KW	DNA-binding; Transcription regulation.			
SQ	SEQUENCE 769 AA; 83774 MW; 73C1FI64FA3F017 CRC64;			

Query Match	28.2%; Score 1062; DB 11; Length 769;
Best Local Similarity	34.9%; Pred. No. 4.9e-73;
Matches 275; Conservative 121; Mismatches 274; Indels 118; Gaps 23;	

Oy	3	PGARGGALARARGGL-----ALLNAVAPLRQAEEIGDGCGHL	44
Dd	13	PGRGCPAAPAATGTAALPSAGCCPLPPRGNSSSRPLLILLIDQDAAGQGDBCGHT	72
Oy	45	VTYODSGTMISKNPYPNHVNCERTIYVPGKRLLIRLGDIIE-SQTCAASYDL-LFT	102
Dd	73	VLGPRSGTLTISINYDHTVNSTVCMEIKIVRGERIRIKFGDFIEDSDYCHLVNTAKYN	132
Oy	103	S---SSDOYGPCG-SMIVPKELLINTSVTVYRFESGHSISRGLITLYAASDHPLTLC	158
Dd	133	GIGVRTEIGKCGAGLNOSIESKGSEVTYLFLMSGTTAAARGFLASTVDKKDLTC	192
Oy	159	LERASHYLETERSKFCPCAGCRDYADVADISGMWDGYRDTSLCKAAITHAIIDELEGQS	218
Dd	193	LDIVSNFLEPEREKCPACCLLPRIAEISTITIHGTIRDSSLPCMAGIGHAGVSNVLGGQS	252
Oy	219	VLARRGISRYEGILANGVLNRDGSLSDKRFLETSMGCSSLSFE----PDGOIRASSSQW	274

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed-11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT *ESDN, a Novel Neupilin-like Membrane Protein Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury.*
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1. SIMILARITY: BELONGS TO THE CRP/FMR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL: AF387547; AAL30178.1; -;
 DR InterPro: IPR000859; JCU domain.
 DR InterPro: IPR000421; PA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR DNA-binding; Transcription regulation.
 KM SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;
 SQ
 Query Match 27.6%; Score 1042; DB 4; Length 775;
 Best local Similarity 36.4%; Pred. No. 1.7e-71;
 Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;
 OY 20 LALLAVSAPRLQAEELDCGCHLTYDSDGTMSTKNTPGYVPHNTVCERTTVPRGKR 79
 DB 51 LLLLVLLLEBDAQOQDGGCHTVLGPESGTLTINPOTYPSNTVCWEWEIRVMKGER 110
 OY 80 LILRLDLDIE-SQTCASDYLFTS----SSDQYGYCG-SMTVPRELLNTSEVTARE 133
 DB 111 VAKREDFEIDSDSCHFYRLRYNGIGVSRTEIGKICGLQAMHSIESKNEITLFLM 170
 OY 134 SSSHSIGRGFLTYASSDHPDLITCLERASHYIKTEYSKFCPCADVADISGNMVDY 193
 DB 171 SGHVSGRGFLASYVIDKODLITCLDTASNFLEPEFSKPCAGCLLPFAEISGTLPHGY 230
 OY 194 ROTSLLCKAIIHAGITADELGGQISYLQKGISRYEGILANGVLSRDGSLDRFLFTSN 253
 DB 231 ROSSPCLMAGVHAGVSNLTGQISVYISKGIPIYESSLANNTSVVGHLSLSTLFETKS 290
 OY 254 GCSRSLSFE---PDQIRASSWSQSVNESGDQVHNSPGQARLQDQPSWASGSSNNHK 309
 DB 291 GCYGTIGMESGYADPQITASSVLENTDHTGQENSRRPKAKLRKKGPPMAAFATDE--- 347
 OY 310 PREMLTIDIGERKKITGITGTSQSNFNYKSFYMANFRKNNNSKKYTKGIVNNEKVF 369
 DB 348 -YQMLDIDLNKREKKTIGITGTSVHEHNYVSAVRIILVSDQOKWTVYREPEVDKIF 406
 OY 370 QGNSNFRDVOVNNFIPPIYARVYVPOTHORIALKVELIGQ-----ITO----- 416
 DB 407 QGNKVDHGVNNFPLPIARFIRVNPQMOQKIAKMLLQGFIPKGRPRKLTQPPPP 466
 OY 417 --GNSLVRKTSQSTSVSTKKEDEITRPI-----PSEET--STGINITT----- 458
 DB 467 RNSND-----LKTITAPPKIAKGRAPKFTQPLQPRSSNEFPATQETTASPDINTVTN 522
 OY 459 ----VAIPVLVLYVFAQMGIFAAE-----RRKKRKS---PYGSAEQKTDCKMQ 503
 DB 523 VTKDVALAALVAVLVAVLTLLILLIVCAMHMRNRKKTGEGYDLPY-----DRAGWMKG 578
 OY 504 IKKYP-----ARQSAEFTISYDNEKEMQKDLISDMADYQOPLMIGTGYT 552
 DB 579 MKQPLPAKAVDHEETPVRSSE--VNHLSPREV--TVLADSLEYQPLGLVIGTL 633
 OY 553 RKGSTFRPMDTAEAG-VSTDAGHYDCPORAGREHYALPLAPPEPATPV---ERH 608
 DB 634 HQRSTFRP--EGGKEAGYADLDP---YNSPGQEVYHAIVAEPLPITQPEVATPIIMDSGH 688

OY 609 VLRAHTFSAGSYRVDPQPGHKNHSLSSGGFSPVAGVADQDGYORPHSAQPADRGYDR 668
 DB 689 PRTSVGQPSSTFRKANGNDP-----PPL--VGYRNTLLSTNDSSSAQADYDR 735
 OY 669 KA 670
 DB 736 KA 737
 RESULT 7
 O8TDX2
 ID O8TDX2 PRELIMINARY; PRT; 775 AA.
 AC O8TDX2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE C1CPL.
 GN C1CPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
 RA Tatematsu Y., Mitsuomi T., Nakao A., Takahashi T.;
 RT *Significant up-regulation of a novel gene, C1CPL, in a highly
 RT metastatic lung cancer subline as well as in lung cancers in vivo.*
 RL Oncogene 0:0-0(2002).
 DR EMBL: AB073146; BAB91138.1; -;
 SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;
 Query Match 27.6%; Score 1042; DB 4; Length 775;
 Best local Similarity 36.4%; Pred. No. 1.7e-71;
 Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;
 OY 20 LALLAVSAPRLQAEELDCGCHLTYDSDGTMSTKNTPGYVPHNTVCERTTVPRGKR 79
 DB 51 LLLLVLLLEBDAQOQDGGCHTVLGPESGTLTINPOTYPSNTVCWEWEIRVMKGER 110
 OY 80 LILRLDLDIE-SQTCASDYLFTS----SSDQYGYCG-SMTVPRELLNTSEVTARE 133
 DB 111 VAKREDFEIDSDSCHFYRLRYNGIGVSRTEIGKICGLQAMHSIESKNEITLFLM 170
 OY 134 SSSHSIGRGFLTYASSDHPDLITCLERASHYIKTEYSKFCPCADVADISGNMVDY 193
 DB 171 SGHVSGRGFLASYVIDKODLITCLDTASNFLEPEFSKPCAGCLLPFAEISGTLPHGY 230
 OY 194 ROTSLLCKAIIHAGITADELGGQISYLQKGISRYEGILANGVLSRDGSLDRFLFTSN 253
 DB 231 ROSSPCLMAGVHAGVSNLTGQISVYISKGIPIYESSLANNTSVVGHLSLSTLFETKS 290
 OY 254 GCSRSLSFE---PDQIRASSWSQSVNESGDQVHNSPGQARLQDQPSWASGSSNNHK 309
 DB 291 GCYGTIGMESGYADPQITASSVLENTDHTGQENSRRPKAKLRKKGPPMAAFATDE--- 347
 OY 310 PREMLTIDIGERKKITGITGTSQSNFNYKSFYMANFRKNNNSKKYTKGIVNNEKVF 369
 DB 348 -YQMLDIDLNKREKKTIGITGTSVHEHNYVSAVRIILVSDQOKWTVYREPEVDKIF 406
 OY 370 QGNSNFRDVOVNNFIPPIYARVYVPOTHORIALKVELIGQ-----ITO----- 416
 DB 407 QGNKVDHGVNNFPLPIARFIRVNPQMOQKIAKMLLQGFIPKGRPRKLTQPPPP 466
 OY 417 --GNSLVRKTSQSTSVSTKKEDEITRPI-----PSEET--STGINITT----- 458
 DB 467 RNSND-----LKTITAPPKIAKGRAPKFTQPLQPRSSNEFPATQETTASPDINTVTN 522
 OY 459 ----VAIPVLVLYVFAQMGIFAAE-----RRKKRKS---PYGSAEQKTDCKMQ 503
 DB 523 VTKDVALAALVAVLVAVLTLLILLIVCAMHMRNRKKTGEGYDLPY-----DRAGWMKG 578

RP SEQUENCE FROM N.A.
RC STRAIN-ILS, AND ISS:
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Gannalf J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Silela J.M.,
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF483506; AAL90780.1; -
DR EMBL: AF483507; AAL90781.1; -
SQ SEQUENCE 926 AA; 104054 MW; 8073D6d78C9A8467 CAC6;

Query Match	9.5%;	Score 357.5;	DB 11;	Length 926;
Best Local Similarity	21.5%;	Pred. No. 1.3e-18;		
Matches 140;	Conservative 97;	Mismatches 241;	Indels 173;	Gaps 22

QY	41	CGHATVYDSDGSMETKRNKPGTYPNHTVEKTTTVC-KGRLLLRIG-DLIDISOTCASPY	98
Db	28	CGGRINSMADAGITSPGTPODYPSPONCENMIYAAEPNOKIYLNPNPHIEIKHCKTYP	87
QY	99	LLF-----TSSSDQVPGYCGSMTPVRELLLNTSEVTVRESGSHISGRGLTYY-----AS	149
Db	88	IEIRGDSESDADLGRKHGNI-APTIISSGSVLTKIFTSVYARQAGASLREIFKTS	146
QY	150	SD-----HPLIIC-----LRASHYLTETYSKF-----	173
Db	147	EDCSKNTSPNGTIESPGFPEKYPHNLCOTFTILAKPRMEIILQELTFPLEHDPLOYVGG	206
QY	174	-CPACRDPVADIS--GNMVDGPRPTSLICKAIAHAGIADDELGOISVLRKGISY--	228
Db	207	DKYMDLWINGIPHVGGLIGTCCTKTPSKRSSTGILSLTFHDMAVAKGFSARYL	266
QY	229	-----EGLANGVLSRDGSLSDKRFLETSNGCSRLSEFPDGOIRAS	270
Db	267	IHOEPENFOCNVPLGAMESGRIANEQISASSPFSQGR-----	303
QY	271	SSMGSVNSSGQVHNSPQOARLQDGPSPWASDSSNNKPREMLTIDGKKKITGIRT	330
Db	304	-----WTPOOSRLHGDNCGMTPNLDN---KELYLDLEFLMTLTAIAIQ	345
QY	331	G--STOSNENFYKSVFVNMFKNNNSKMTYKGIYVNEKEVFGQNSNFRPVOYNNFIPY	388
Db	346	GAISHTOKGYVKSRYKLEVSINGSDMWYRGNKH--KIPANDATPEVYLNKLHMLL	403
QY	389	ARYAVVDPQTHORIALVELGCOITQDGSLSVWRKTS--QSTSVSTKKDETTTPI	445
Db	404	TRFIRIPQTHLGLNLRLELFGCEVTAPCSNMGLMGLIADPQISASTREYLMSPS	463
QY	446	-----PSEESTGINTITVALPIVLIVLYVFAQMGIFA---APRK	482
Db	464	AARLYSSSGMFPNRPQAPQGEW-LQYDLGTPKTVKGITIQGANGGDSITVEARAFAFR	522
QY	483	KKKKSPGSAEAKOTCDCKOIKYFPAHQSA-EFTISD-----	521
Db	523	KFKYSYSLNGAD-----WEYIQDRHQOTKIFEGNMHYDPDIRFDPVPAQYRAYPE	576
QY	522	--NEKEMTKIDLITSDMADYQOPL-MIGGTVTYRKGSTFRPMDTAEAG	569
Db	577	RWSPAGIGMRLEVLGCDMTDSKPTVETIGPVTYKSEETTPYPMDEDATGECG	627
RESULT 11			
Q9H2E4	ID	Q9H2E4	PRELIMINARY; PRT; 901 AA.
AC	Q9H2E4		
DT	01-MAR-2001	(TREMBLrel. 16, Created)	
DT	01-MAR-2001	(TREMBLrel. 16, last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, last annotation update)	
DE	Neutrophilin-2b(O).		
GN	NR2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
PN	[1]		
RP	SEQUENCE FROM N.A.		

RA MEDLINE:20564205; PubMed:11112349;
RX Rossignol M., Gagnon M.L., Klagesbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONSTRAINTS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONSTRAINTS 1 MAM DOMAIN.
DR EMBL: AF280544; AACG1403.1; -.
DR HSSP: P12259; 1CET.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR000998; MAM_domain.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00754; P5_F8_type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SMO0042; CUB; 2.
DR SMART: SMO0231; FA58C; 2..
DR SMART: SMO0137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PSS0060; MAM_2; 1.
KW Glycoprotein; Neurone; Receptor; Transmembrane.
SQ SEQUENCE 901 AA; 101380 MW; 1F3107A9CC65E2A2 CRC64;

Query Match	9.38;	Score 351.5;	DB 4;	Length 901;
Best Local Similarity	22.2%;	Pred. No. 3.6e-18;		
Matches 142;	Conservative 99;	Mismatches 255;	Indels 143;	Gaps 25

```

0Y 41 CSHLTVOGDSGSMKNNPGVTPNHYCEKTTIVK--GKRLLRLG--DLIDSQCASOY 98
   | : : : : : | : | : | : : : : : : : : : : : | : | : | :
Db 28 CGGRLSNDAGITTPGPIPODYPSONCEMIYAPENOKIYLNPNHFEIENHCKYF 87
   | : : : : : | : | : | : : : : : : : : : : : | : | : | :
0Y 99 LLE----TSSDOYPGYCGSMTPRELLANTSEVTYRFEBSGSHISGRGILTY-----AS 149
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 88 IEIRDQSESDADLCKHCGNI--APPTIISGSMLYIKFTSDVARQAGSLNREIPIKTS 146
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
0Y 150 SD-----HPLDITC-----LEBASLYLKEYSKF----- 173
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 147 EDCSKNFTSPNGTIESPGFPERKYPHNLDCTFTILAKPKMEILLOFLIDPLEHDPLOVGE 206
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 174 -CPACRQVADIS--GNWVDGYRDTSLCKKAIHAGIITADELGGISVLQKRGSRYS 230
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 207 DCKYIMLDMIDGIPHYGLIGKTCGTKTPSELSRSTGIISLTFPHDMAAK-----DG 259
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 231 ILANGVLSDGSLSDKRELTSGCSRSLSFE-----PDQIARASSWQSVNESGDQVHS 286
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 260 FSARYIYVHQELENFQ-----CNVPLGMSGRILANQIASSTY-----SDGRMT 305
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 287 PGQARLQDQGSWASGSDSSNNHKKPREMLEIDGEEKKITGIRTTG--STQSNFNEYKSF 344
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 306 FOOSRLHDDNDGMPNLDNSN---KEYLOVDRFLTYMLTALITQOASIRETONGYVYKSY 361
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 345 VMNFNKNSKMYTGCIYVNNKEVYOGNSNFDPQUNNPIPIVARYVYVYVOTHQRIA 404
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 362 KLEVTNSEDMMVYTHGKNH--KVYQANNDATLEVYLNKILHAPLITREVIPIRQTHWSGTA 419
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 405 LKVELLIGCOITOGNDSLWVRKTS---OSTSVSTKKEDETITRP-----IP 446
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 420 LRLLEFGCRVYDAPCSNMLGMLSGILADSOIASASTOETLWBSARLYSNSGWEPRIP 479
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 447 SEENSTGINITTVAILPYLVLLVYLVFAGMG---IFA---AFKKKKKSGSPYSAEAOKTD 499
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 480 QAOPEEMLOVDLGPKKYVKVYIIOGARGGDSITVEARAARFKKFKVSYSLNGKD----- 534
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
0Y 500 CMKOIKYPAFHQSA--ETIISYD-----NEKMTOKDLITSD 536
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 535 -WEYIODRTOOPKLEEGNMHYDPDIRFDDIPAOYVARYVPERKSPAGIGMLTGLGD 593
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

```

QY 537 MADYQPLMIGTGYTRKGFTRPMDTDAEAGVSTDAG 575
DB 594 WTD-SKPTVETLGPVKSEETTPPTTEE-----ATECG 627

RESULT 12
09H2D5 PRELIMINARY; PRT; 901 AA.
AC 09H2D5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Neuropilin-2b(0).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF281074; AAC41899.1; -.
DR HSSP: P12259; 1C2T.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00442; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS50060; MAM_2; 1.
DR GlycoProfile: Neutrone; Receptor; Transmembrane.
SQ SEQUENCE 901 AA; 101381 MW; A1310304AFD086AB CRC64;

Query Match 9.3%; Score 351.5; DB 4; Length 901;
Best Local Similarity 22.2%; Pred. No. 3.6e-18;
Matches 142; Conservative 99; Mismatches 25; Indels 143; Gaps 25;

QY 41 CGHVTYDGSMTSKNPGYVPHNTVEKTIYVK-CKRLILRLG-DLDESQCASDY 98
DB 28 CGGRNLSMDAGIYISPGYPODYPSHONCEMYIAPERPCKIYLVNPNPELEKHKCKYDF 87
QY 99 LIF-----TSSSDQGYPGYSGMTVPKELLNLTSEVTVRESGSHISGRFLTY-----AS 149
DB 88 IEIRGDSESDLLCKHCGNI-APPTIISGSMYIKFTSDVARGAGFSLRYEIFKGS 146
QY 150 SD-----HDLITC-----LERASHYLKTEYSK----- 173
DB 147 EDCSKNFTSPNGITSPGPFPERKPHNLDCTFTILAKPKMEIILDFLIDLEHDPVQVGE 206
QY 174 -CPAGCRVADIS--GNMVDGYRFTSLCKAIAHAGIITADELGGQIVLVLRKGSRYEG 230
DB 207 DCKIDMDIWMGIRAVGVLIGKYGCTKTPSELRSSTGILSLFPHDMAVAK-----DG 259
QY 231 ILANGVLSRDSLSDKRFLFTSNGCSRLSFE---PDGQIRASSSQVSNESGQDVHMS 286
DB 260 FSARYLVVHQBLENFQ-----CNVPLGMSGRIANEQISASTY-----SDGKWT 305

```

```

QY 287 PQQARLDQOGSGWASGDSNNHKKPREMLEIDLGKKKITGTITTC--STQSNFNFYKSF 344
DB 306 PQOSRLHGDNDGNWPTNLSN-----KEYLDVLDLFTLMTLALATQGAISRETONGYVKSX 361
QY 345 VNFNKNNSKMTKYGIYVNEEKYVQGSNFPDVPQNHFIPYVARYVYVPOTHQRIA 404
DB 362 KLEVSTNGEDMKVYTRHGNH--KVFQANNDATEVVLNLAHLRFRVIRNPQTHSGIA 419
QY 405 LKVELIGQITQGNDSLVAWRKTS---QSTSVSTKKEDETITRP-----IP 446
DB 420 LRLEIFGCRVTDAPCSNMGLSLIADQSISASTQGYLMSPSARLVSSRSGWFRIP 479
QY 447 SEETSTGINITTVAILPVLLVLYVFAAG---IFA-----ARKKKKKSPYGSAPQAKTD 499
DB 480 QAQPEEMLYQVLDGTPKTVKGVIIQAGRGDSITAVEARAFVRFKYSYSLNGKD----- 534
QY 500 CWKQIKYPEFARHQA-EETISYD-----NEKEMTKDLINSD 536
DB 535 -WEYIDPRTPQPKLFEENMHYDTPDIRFDPIPAQYVRYVPERKNSPAGICMRLEVLCD 593
QY 537 MADYQPLMIGTGYTRKGFTRPMDTDAEAGVSTDAG 575
DB 594 WTD-SKPTVETLGPVKSEETTPPTTEE-----ATECG 627

RESULT 13
09H2E3 PRELIMINARY; PRT; 906 AA.
AC 09H2E3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Neuropilin-2b(5).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF280545; AAC41404.1; -.
DR HSSP: P12259; 1C2T.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00442; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS50060; MAM_2; 1.
DR GlycoProfile: Neutrone; Receptor; Transmembrane.
SQ SEQUENCE 906 AA; 101955 MW; 0FB01F24DBD9B0BD CRC64;

Query Match 9.3%; Score 351.5; DB 4; Length 906;
Best Local Similarity 22.2%; Pred. No. 3.7e-18;
Matches 142; Conservative 99; Mismatches 25; Indels 143; Gaps 25;

```


OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Herzog Y., Kalchauer C., Kahane N., Reshef R., Neufeld G.;
 RT "Differential expression of neuropilin-1 and neuropilin-2 in arteries
 and veins."
 RL Mech. Dev. 0:0-0(2002)
 DR EMBL; AF417235; AA37723.1; -;
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00754; PS_F8_type_C; 2.
 DR Pfam; PF00629; MAM; 1.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; UNKNOWN_2.
 DR PROSITE; PS01286; FA58C_2; UNKNOWN_2.
 DR PROSITE; PS50060; MAM_2; 1.
 DR Receptor.
 KW Receptor.
 SQ SEQUENCE 919 AA; 103154 MW; 418AB528D62D59B4 CRC64;

Query Match 9.38; Score 351; DB 13; Length 919;
 Best Local Similarity 24.98; Pred. No. 4.1e-18;
 Matches 115; Conservative 71; Mismatches 173; Indels 102; Gaps 18;

QY 18 GLALLLAVSAPRLQAEELDGCGHLVTYDSGTMTSKNYGTYPNHTVCEKTTVPK- 76
 DB 9 GFLGCLPGSG---GAETANQPGGRINSKDAGYITSPGYPNDSHQNCEWYIAPEP 64
 QY 77 GKRLILRG-DLDIESQTCADYLLF---TSSDQGYPCGSKMTPKELLNLTSEYTVR 131
 DB 65 NOKIILNPNPFEIEKHCKYDFEIRDDSEADLGRKHCNI-APPTIISGSLYIK 123
 QY 132 FESGSHISGRGFLTY-----ASSD-----HPDLITCLERASHYL 166
 DB 124 FTSDIAROGAGFSLRYEYKTSSEDCSNFTASNGTIESPGFPDKYPHNLCVFTIITAKP 183
 QY 167 KTE---YSKE-----CPAGCRDY-----AGDISG-----NMVDGYRDTSL 198
 DB 184 KTEILLHFVLFDELHPDLPQAGEGCKYDMLDIPQVGPLIGYCGTKMPSDIRSTTG 243
 QY 199 LCKRAIIR--GIIDELGGQISVLQKGISRYEGILANGVLSRSGSLSDKRFITSNCS 256
 DB 244 VLSLTFHTDLAVARDGFSQYLLIQOEVPEFQCNPVIGMES--GRISNM----- 291
 QY 257 RSLSEPDGQIRASSSMQSVNESGDQVHWSPGQARLDQGPSMASGSSNNHKPREMLEI 316
 DB 292 -----QISASSTY-----SDGKMTPOQSRHNSDDNGWTPNVDN-----KEYLQV 331
 QY 317 DLGKKKKTIGIRTG--STGSNFNFYKSFVYNNFNKNNNSKNTYKGIYVNEKEVQGSNN 374
 DB 332 DLHFLVLTALATOGAISRFTONGYVRYTYLKVSTNGEDMMYRHKNH--KTFQANED 389
 QY 375 FRDPVQNNFIPIYARVRYVPQTHQRIAKVELIGCQIT 415
 DB 390 ATEVVLNKHISPVLTFRVIRIPQSHNGIALRLLEYGCRI 430

Search completed: May 15, 2003, 13:18:46
 Job time : 75.8563 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 64.5731 Seconds

(without alignments)
1475.448 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGARCGALARAARGRL.....TSAPRDLTPINOTAMTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq 101002:*

- 1: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3770	100.0	715	22	AAU00670 Human TANGO 229 po
2	2812	74.6	539	22	AAU00630 Novel human protei
3	2812	74.6	586	22	AAU00629 Novel human protei
4	2551	67.7	487	22	AAU00628 Novel human protei
5	1912.5	50.7	503	21	AAAB19126 Polypeptide isolat
6	1409.5	37.4	398	23	AAE22715 Human neuropilin-H
7	1409.5	37.4	398	23	AAU79459 Human neuropilin-H
8	904	24.0	669	21	AAV70539 Human Factor 8 Hom
9	902	23.9	583	21	AAAG75450 Human colon cancer
10	692.5	18.4	385	23	AAE22716 Human neuropilin-H

11	692.5	18.4	385	23	AAU79460	Human Neuropilin-H
12	692.5	18.4	385	23	ABE97386	Novel human protei
13	684.5	18.2	365	23	AAE22721	Human neuropilin-H
14	364.5	9.7	889	21	AAAB24216	Soluble neuropilin
15	362.5	9.6	925	20	AAW96308	Neuropilin-2, Rat
16	354.5	9.4	901	20	AAW96256	Mouse semaphorin r
17	354.5	9.4	906	20	AAW96257	Mouse semaphorin r
18	354.5	9.4	909	20	AAW96249	Rat semaphorin rec
19	354.5	9.4	909	20	AAW96250	Mouse semaphorin r
20	354.5	9.4	909	21	AAAB24214	Mouse soluble neur
21	354.5	9.4	914	20	AAW96251	Mouse semaphorin r
22	354.5	9.4	926	20	AAW96252	Mouse semaphorin r
23	354.5	9.4	931	20	AAW96253	Mouse semaphorin r
24	351.5	9.3	909	20	AAW96254	Mouse semaphorin r
25	351.5	9.3	926	20	AAW96255	Human semaphorin r
26	351.5	9.3	926	23	AAE22718	Human neuropilin-2
27	351.5	9.3	931	20	AAV06318	Human neuropilin-2
28	351.5	9.3	931	20	AAV14563	Human VEGF(165)/N
29	351.5	9.3	931	20	AAV23248	Human NP-2 protein
30	349	9.3	555	22	AAAB64627	Human secreted pro
31	349	9.3	555	22	AAW06320	Human soluble neur
32	345.5	9.2	487	22	AAAB64570	Human secreted pro
33	341	9.0	439	22	AAAB64628	Human secreted pro
34	338.5	9.0	892	21	AAAB24215	Soluble neuropilin
35	338.5	9.0	923	20	AAW96248	Mouse semaphorin r
36	338.5	9.0	923	21	AAAB24213	Mouse soluble neur
37	338.5	9.0	923	23	ABE57098	Mouse ischaemic co
38	335	8.9	538	22	AAU02950	Angiotensin conver
39	335	8.9	600	22	AAU02948	Angiotensin conver
40	335	8.9	644	20	AAV06319	Human soluble neur
41	335	8.9	840	22	AAU02949	Angiotensin conver
42	335	8.9	856	22	AAAB62478	Human NP-1 recepto
43	335	8.9	923	20	AAV06317	Human VEGF(165)/N
44	335	8.9	923	20	AAV23247	Human neuropilin-1
45	335	8.9	923	20	AAW96246	Human semaphorin r

ALIGNMENTS

RESULT 1	AAU00670	AAU00670 standard; Protein; 715 AA.
ID	AAU00670	
XX	AAU00670;	
AC	07-SEP-2001 (first entry)	
XX		
DT	07-SEP-2001 (first entry)	
XX		
DE	Human TANGO 229 polypeptide.	
XX		
KW	Human: TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;	
KW	skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;	
KW	bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;	
KW	anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;	
KW	malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;	
KW	attention deficit disorder; Crohn's disease; gastroenteritis; goutte;	
KW	hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;	
KW	muscular dystrophy; immuno-competence; vertebrate; blood; serum.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
Peptide	1..34	
FT	/note="Signal peptide"	
FT	35..455	
FT	/note="Extracellular domain"	
FT	35..715	
FT	/note="Mature human TANGO 229"	
FT	456..480	
FT	/note="Transmembrane domain"	
FT	481..715	
FT	/note="Cytoplasmic domain"	
XX		

PN W0200129088-A1.
 XX 26-APR-2001.
 PD 23-JUN-2000; 2000WO-US17386.
 PF 19-OCT-1999; 99US-0420707.
 XX
 PR
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 PI Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leahy KR;
 XX
 DR WPI: 2001-308477/32.
 DR N-PSDB; AAS00660.
 XX
 PT New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes -
 XX
 PS Claim 8; Fig 1; 263pp; English.
 XX
 CC The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration
 CC of, e.g., T cells and cells of the heart, liver, pancreas, placenta,
 CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
 CC blood leukocyte, bone marrow or thymus tissue. They can be used as
 CC modulating agents for regulating cellular processes, thus, the proteins
 CC and their associated nucleic acids can be used to prognosticate, prevent,
 CC diagnose, or treat disorders associated with physiological processes.
 CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
 CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
 CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
 CC deficit disorder, Crohn's disease, gastroenteritis, golfire,
 CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
 CC muscular dystrophy. Antibodies to disorders such as these can be made by
 CC providing a polypeptide of the invention to an immuno-competent
 CC vertebrate and harvesting blood or serum from the vertebrate.
 CC
 XX
 SQ Sequence 715 AA;
 Query Match 100.0%; Score 3770; DB 22; Length 715;
 Best Local Similarity 100.0%; Pred. No. 1e-306;
 Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYPGARGGALARAAGRLALLAVSAPRLQAELGCGHLYTYDSCGTMSTKNPG 60
 DB 1 MYPGARGGALARAAGRLALLAVSAPRLQAELGCGHLYTYDSCGTMSTKNPG 60
 QY 61 TYPNHTVEKITYPKGRILIRGLDIESQTCASDYLFTSSSDOYGPYCGSMYVKE 120
 DB 61 TYPNHTVEKITYPKGRILIRGLDIESQTCASDYLFTSSSDOYGPYCGSMYVKE 120
 QY 121 LLMTSEVTTFESGSHISGRFLITYASSDHPDITCLEASHLTKYEYKFCPAGGRD 180
 DB 121 LLMTSEVTTFESGSHISGRFLITYASSDHPDITCLEASHLTKYEYKFCPAGGRD 180
 QY 181 VAGDISGNMVDGYRDTSLICAAIHAGIADLGGQISVLQKGISRYEGILANGVLSRD 240
 DB 181 VAGDISGNMVDGYRDTSLICAAIHAGIADLGGQISVLQKGISRYEGILANGVLSRD 240
 QY 241 GSLDKRILFTSNCGSRSLSEFPDQIRASSWOSVNSGQVHMSPOQARLDQGPMSMA 300
 DB 241 GSLDKRILFTSNCGSRSLSEFPDQIRASSWOSVNSGQVHMSPOQARLDQGPMSMA 300
 QY 301 SGDSNNHKKPREMLEIDGEEKKITGIRTTGOSTQSNFNYKSFYMNKNNNSKMYTKG 360
 DB 301 SGDSNNHKKPREMLEIDGEEKKITGIRTTGOSTQSNFNYKSFYMNKNNNSKMYTKG 360
 QY 361 IVNNEKVFQGSNFRDPVQNNFIPPIYARVRYVPQTHORIALKVELICQITQGNDS 420
 DB 361 IVNNEKVFQGSNFRDPVQNNFIPPIYARVRYVPQTHORIALKVELICQITQGNDS 420

QY 421 LVWRKTSOSTSVSTKKEDEITTRPIPSEETSTGINITTVAILVLLVLYVPAKGCIFAPAF 480
 DB 421 LVWRKTSOSTSVSTKKEDEITTRPIPSEETSTGINITTVAILVLLVLYVPAKGCIFAPAF 480
 QY 481 RRRKKKSPYSGAEAKQDKCKQKIRYPPARHQSAEFTTSYNEKEMOKLTLTSDMADY 540
 DB 481 RRRKKKSPYSGAEAKQDKCKQKIRYPPARHQSAEFTTSYNEKEMOKLTLTSDMADY 540
 QY 541 QOPLMIGTGVTRKSGTFRPMDTDAEAGVSTDAGHYDCPORAGRHEALPLAPPEPEY 600
 DB 541 QOPLMIGTGVTRKSGTFRPMDTDAEAGVSTDAGHYDCPORAGRHEALPLAPPEPEY 600
 QY 601 ATPYERHVLAAHFFSAQSGTRVPGPOPKHKSJSSGGSFVYAGVGAODGYQRPHSQAP 660
 DB 601 ATPYERHVLAAHFFSAQSGTRVPGPOPKHKSJSSGGSFVYAGVGAODGYQRPHSQAP 660
 QY 661 ADRGDRPKAVSALATESGHPDSQKPPHPSTSPSYAPROCLTLPNOTAMTALL 715
 DB 661 ADRGDRPKAVSALATESGHPDSQKPPHPSTSPSYAPROCLTLPNOTAMTALL 715
 RESULT 2
 ID A000630 standard; Protein: 539 AA.
 AC A000630;
 AC A000630;
 DT 29-APR-2001 (first entry)
 XX
 DE Novel human protein (NHP) sequence #3.
 XX
 KW Novel human protein: NHP; CUB domain; extracellular domain; gene therapy;
 KW obesity; high blood pressure; connective tissue disorder; infertility;
 XX NHP-mediated pathway.
 OS Homo sapiens.
 XX
 PN W0200129219-A1.
 PD 26-APR-2001.
 PF 08-OCT-2000; 2000WO-US28798.
 XX
 PR 19-OCT-1999; 99US-0160285.
 XX
 PR 18-FEB-2000; 2000US-0183583.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX
 DR WPI: 2001-290917/30.
 DR N-PSDB; AAS00615.
 XX
 PT Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX
 PS Claim 5; Page 29-30; 33pp; English.
 XX
 CC The sequence represents a novel human protein (NHP) containing a CUB
 CC domain (an extracellular domain). CUB proteins have been associated with
 CC regulating development, modulating cellular processes and preventing
 CC infectious disease. NHP nucleotide sequences are useful for gene therapy
 CC of physiological disorders or diseases. NHP oligonucleotides are useful
 CC as hybridisation probes for screening libraries and assessing gene
 CC patterns. NHP nucleotide sequences are useful for detecting mutant or
 CC inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These

CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

SQ Sequence 539 AA;

Query Match 74.6%; Score 2812; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.2e-226;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MYPGAGGALARAAGRLATALLAVSAPRLQAEELGDCGHLVYODSGTMSKNYPC 60
DB 1 MYPGAGGALARAAGRLATALLAVSAPRLQAEELGDCGHLVYODSGTMSKNYPC 60
OY 61 TYPNHTVCEKTIIVPKGRKRLRLGLDLIESQTCASDYLLFTSSPOYGYCGSMVYPK 120
DB 61 TYPNHTVCEKTIIVPKGRKRLRLGLDLIESQTCASDYLLFTSSPOYGYCGSMVYPK 120
OY 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSKFCPCAGCRD 180
DB 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSKFCPCAGCRD 180
OY 181 VAGDISGNMVDGYRDTSLCKAIIHAGITADELGGQISVLRKGISRYEGITLANGVLSRD 240
DB 181 VAGDISGNMVDGYRDTSLCKAIIHAGITADELGGQISVLRKGISRYEGITLANGVLSRD 240
OY 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQSVNESGDVHWSPGQARLDQGPSPA 300
DB 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQSVNESGDVHWSPGQARLDQGPSPA 300
OY 301 SGDSNNHNPREFLEIDLEKKKITGIRTTGTSQSNFNFYVKSFWANFKNNKMKTYGK 360
DB 301 SGDSNNHNPREFLEIDLEKKKITGIRTTGTSQSNFNFYVKSFWANFKNNKMKTYGK 360
OY 361 IYNNKEKVEQGSNENFDPYONNFIPIVARYVVPQTHORALVELIGCOITGNGDS 420
DB 361 IYNNKEKVEQGSNENFDPYONNFIPIVARYVVPQTHORALVELIGCOITGNGDS 420
OY 421 LVWRKTSOSTSVSTKKEDEITRPIPESETSTGINITVAIPLVLVLFAGMGJFAAF 480
DB 421 LVWRKTSOSTSVSTKKEDEITRPIPESETSTGINITVAIPLVLVLFAGMGJFAAF 480
OY 481 RKKKKGSPYGSAAQKTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLDLITSMA 538
DB 481 RKKKKGSPYGSAAQKTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLDLITSMA 538

```

RESULT 3

AAU00629 standard; Protein; 586 AA.

```

ID AAU00629 standard; Protein; 586 AA.
XX
AC AAU00629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
XX
PA 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX

```

PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00614.
XX

PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 2; Page 27-28; 33pp; English.

XX The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridization probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

SQ Sequence 586 AA;

Query Match 74.6%; Score 2812; DB 22; Length 586;
Best Local Similarity 100.0%; Pred. No. 1.4e-226;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MYPGAGGALARAAGRLATALLAVSAPRLQAEELGDCGHLVYODSGTMSKNYPC 60
DB 48 MYPGAGGALARAAGRLATALLAVSAPRLQAEELGDCGHLVYODSGTMSKNYPC 107
OY 61 TYPNHTVCEKTIIVPKGRKRLRLGLDLIESQTCASDYLLFTSSPOYGYCGSMVYPK 120
DB 108 TYPNHTVCEKTIIVPKGRKRLRLGLDLIESQTCASDYLLFTSSPOYGYCGSMVYPK 167
OY 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSKFCPCAGCRD 180
DB 168 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSKFCPCAGCRD 227
OY 181 VAGDISGNMVDGYRDTSLCKAIIHAGITADELGGQISVLRKGISRYEGITLANGVLSRD 240
DB 228 VAGDISGNMVDGYRDTSLCKAIIHAGITADELGGQISVLRKGISRYEGITLANGVLSRD 287
OY 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQSVNESGDVHWSPGQARLDQGPSPA 300
DB 288 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQSVNESGDVHWSPGQARLDQGPSPA 347
OY 301 SGDSNNHNPREFLEIDLEKKKITGIRTTGTSQSNFNFYVKSFWANFKNNKMKTYGK 360
DB 348 SGDSNNHNPREFLEIDLEKKKITGIRTTGTSQSNFNFYVKSFWANFKNNKMKTYGK 407
OY 361 IYNNKEKVEQGSNENFDPYONNFIPIVARYVVPQTHORALVELIGCOITGNGDS 420
DB 408 IYNNKEKVEQGSNENFDPYONNFIPIVARYVVPQTHORALVELIGCOITGNGDS 467
OY 421 LVWRKTSOSTSVSTKKEDEITRPIPESETSTGINITVAIPLVLVLFAGMGJFAAF 480
DB 468 LVWRKTSOSTSVSTKKEDEITRPIPESETSTGINITVAIPLVLVLFAGMGJFAAF 527
OY 481 RKKKKGSPYGSAAQKTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLDLITSMA 538
DB 528 RKKKKGSPYGSAAQKTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLDLITSMA 585

```

RESULT 4

AAU00628

ID AA00628 standard; Protein: 487 AA.
XX
AC AA00628;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #1.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-0528798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
PI WPI: 2001-290917/30.
DR N-PSDB; AAS00613.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 4; Page 26; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
SO Sequence 487 AA;
Query Match 67.7%; Score 2551; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 7.9e-205;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 53 MTSNYPGTYNHRVCEETITVPKGRLLRLGDDIESQTCASDYLTFTSSDDYGYC 112
DB 1 MTSNYPGTYNHRVCEETITVPKGRLLRLGDDIESQTCASDYLTFTSSDDYGYC 60
OY 113 GSMVPRKLLNTSEVYRFESGSHISGRGLLYVASSDHDLLTCLERASHYLTETSK 172
DB 61 GSMVPRKLLNTSEVYRFESGSHISGRGLLYVASSDHDLLTCLERASHYLTETSK 120
OY 173 FCPAGCNDVADISGNMVDGYRDTSLCKAIIHAGIINDELJGQISVLRKRGISRYEGIL 232
DB 121 FCPAGCNDVADISGNMVDGYRDTSLCKAIIHAGIINDELJGQISVLRKRGISRYEGIL 180
OY 233 ANGLSRGSLSDKRFLFTSGCSRSLSFEPDQIRASSSQSVNESDQVHMSGQARL 292
|||||

DB 181 ANGLSRGSLSDKRFLFTSGCSRSLSFEPDQIRASSSQSVNESDQVHMSGQARL 240
OY 293 ODQGPASGDSNNHKKREMLEIDLGKKKKTGIRTGSTOSNPNFYVKSFPVANKNN 352
DB 241 ODQGPASGDSNNHKKREMLEIDLGKKKKTGIRTGSTOSNPNFYVKSFPVANKNN 300
OY 353 SKWKTYKIVNNEKEVFQGSNFRDPVONNFIPIVARYVAVPOTVHRIALKEVIGC 412
DB 301 SKWKTYKIVNNEKEVFQGSNFRDPVONNFIPIVARYVAVPOTVHRIALKEVIGC 360
OY 413 QITGNDSLVWRKTSQSSTVSTKKEDETITRPISEESTSGINTTVAIPVLVLYEA 472
DB 361 QITGNDSLVWRKTSQSSTVSTKKEDETITRPISEESTSGINTTVAIPVLVLYEA 420
OY 473 GNGJFAFRKKKKKSGPYGSAEAKTDCWKQIKYPFARHQAEPFISYDNKEWTKIDL 532
DB 421 GNGJFAFRKKKKKSGPYGSAEAKTDCWKQIKYPFARHQAEPFISYDNKEWTKIDL 480
OY 533 ITSDMA 538
DB 481 ITSDMA 486
RESULT 5
AAB19126
ID AAB19126 standard; Protein: 503 AA.
XX
AC AAB19126;
XX
DT 19-FEB-2001 (first entry)
XX
DE Polypeptide isolated from lymph node stromal cells of fsn -/- mice.
XX
KW Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW immune system disorder; cancer; viral infection; HIV infection;
KW blood vessel growth; tumour necrosis factor disorder; arthritis;
KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW cardiac failure.
XX
OS Mus sp.
XX
PN WO2000058463-A1.
XX
PD 05-OCT-2000.
XX
PR 18-FEB-2000; 2000WO-NZ00015.
PR 25-MAR-1999; 99US-0276268.
PR 26-AUG-1999; 99US-0383586.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PI Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI Murlison JG;
XX
DR WPI: 2000-664924/64.
DR N-PSDB; AAA96736.
XX
PT Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
PT useful for modulating growth of blood cells, for treating inflammatory
PT and tumour necrosis factor-mediated disorders, cancer and viral
PT disorders -
XX
PS Claim 1; Page 68-69; 75pp; English.
XX
CC The present sequence represents a polypeptide sequence which is
CC isolated from lymph node stromal cells of fsn -/- mice. The
CC polynucleotides and their polypeptides are useful for treating an
CC inflammatory disorder, disorder of immune system and cancer selected
CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC viral disorder, in particular HIV infection and for modulating the
CC growth of blood vessels. The polypeptides are useful for treating a
CC tumour necrosis factor (TNF) mediated disorder, such as those selected

CC from arthritis, inflammatory bowel disease and cardiac failure and a
 CC fibroblast growth factor-mediated disorder. It is also useful in assays
 CC to determine biological activity, to raise antibodies, to isolate
 CC corresponding ligands or receptors, to quantify levels of protein or
 CC cognate corresponding ligand or receptors, as anti-inflammatory agents,
 CC and in compositions for the treatment of skin, connective tissue and
 CC immune system diseases. The polynucleotide is useful as marker for
 CC tissue, as a chromosome marker or tags in the identification of a
 CC genetic disorder.

XX Sequence 503 AA;

Query Match 50.7%; Score 1912.5; DB 21; Length 503;
 Best Local Similarity 55.6%; Pred. No. 2,6e-151;
 Matches 390; Conservative 35; Mismatches 74; Indels 203; Gaps 4;

QY 14 AAGGIALALLAVSAPRLQAEELGDCGHLVYQDSGMTSKRYPCGTYPHNTCECTIT 73
 DB 5 AGGVSVALPNCAPRLQAEELGDCGHLVYQDSGMTSKRYPCGTYPHNTCECTIT 64
 QY 74 VPKGRLLILRLGLDIESQTCASDYLLFTSSDQYGCYSMTVPKELLNTSEVTRFE 133
 DB 65 VPKGRLLILRLGLDIESQTCASDYLLFTSSDQYGCYSMTVPKELLNTSEVTRFE 124
 QY 134 SGSHISGRGLLYASSDHPDLITCLERASHYLTKEYSKFCPCACRDVADISGNMVDGY 193
 DB 125 SGSHISGRGLLYASSDHPDLITCLERASHYLTKEYSKFCPCACRDVADISGNMVDGY 184
 QY 194 RDSILCKAIIHGIINDELGGQISVLRKISREGLIANGVLSRDSISDRKFLTSN 253
 DB 185 RDSILCKAIIHGIINDELGGQISVLRKISREGLIANGVLSRDSISDRKFLTSN 241
 QY 254 GCSNLSFEPDGOIRASSSQSVNESGDQYHMSFGARLDQGPSMASGDSNNHREPW 313
 DB 242 ----- 241
 QY 314 LEIDLGKKKKTIGRTGTSQSNFNYKSFVNNKNNNSKMKTYKGIYNNKEKVFQONS 373
 DB 242 ----- 241
 QY 374 NFRDPQNNFPIPIVAVYVVPQTWHQRIALNVELIGCQITQGNDSLVWRKTSQSTVS 433
 DB 242 ----- 241
 QY 434 TKREDETTPRIPISEETSTGINTTVAIPLVLVVFAGMGIFAPRKKKKKSPYGS 493
 DB 242 ----- 241
 QY 494 EAQKTDCKWKQIKTPPARHQSAPETISYDNEMKQKLDLITSDMADYQOPLMIGTVTR 553
 DB 286 DAQKTGCKWKQIKTPPARHQSAPETISYDNEMKQKLDLITSDMADYQOPLMIGTVTR 345
 QY 554 KGSFPRMDADAEAGYSTAGGHYDCROBAGREVALPLAPPEPVAATPIVERHVRAN 613
 DB 346 KGSFPRMDADAEAGYSTAGGHYDCROBAGREVALPLAPPEPVAATPIVERHVRAN 405
 QY 614 TFSNAGSYRVPGQPGHKSLSGSGFSPVAGVAGQDQYORPHASQPADRGYDRPKAVSA 673
 DB 406 TFSNAGSYRVPGQPGHKSLSGSGFSPVAGVAGQDQYORPHASQPADRGYDRPKAVSA 464
 QY 674 LATESGHPDSQKPTNAGTSDSISAPDCITPLNQTMTALL 715
 DB 465 L-DSRDPASQSQMT-SGGDGYSAFPRNGLAFLNQTMTALL 503

RESULT 6
 AAE22715
 ID AAE22715 standard; Protein; 398 AA.

XX AAE22715;
 AC
 XX
 DT 09-AUG-2002 (first entry)
 XX

DE Human neuropilin-Hy1 protein.

XX Human: neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnerability; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; viroicide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.

XX Homo sapiens.

XX WO200222815-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28488.

XX 11-SEP-2001; 2000US-0659671.

XX 06-SEP-2001; 2001US-317902P.

XX (HYSE-) HYSEQ INC.

XX Tang YT;

XX WPI: 2002-393966/42.

XX N-PSDB; AAD35992.

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects

XX Claim 3; Page 123-125; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy1
 CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

XX Sequence 398 AA;

Query Match 37.4%; Score 1409.5; DB 23; Length 398;
 Best Local Similarity 72.8%; Pred. No. 2,6e-109;

Matches 289; Conservative 11; Mismatches 26; Indels 71; Gaps 7;

OY 38 GDGCGHLYTYDSCGTMSTKNTPGTYPNHTVCEKTTVPKGRKRLIRLGDLDIESQTCASD 97
 DB 4 GDGCGHLYTYDSCGTMSTKNTPGTYPNHTVCEKTTVPKGRKRLIRLGDLDIESQTCASD 63
 OY 98 YLFTSSSDQGYGCGSMTPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDILT 157
 DB 64 YLFTSSSDQGYGCGSMTPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDILT 121
 OY 158 CLERSHLYKTEYSK-FCPACGRDYAGDISGMNDGYRDTSLCKRAIHAGIADDELGGQ 216
 DB 122 QGDRSEKTLDOQSRFTLATGTFVKDSFS--TDG---TSLCKRAIHAGIADDELGGQ 175
 OY 217 ISVLORKGISREGLANGVLS-----FLFTSNGCSRSLSEFPDGOIRASSSMOS 238
 DB 176 ISVLORKGISREGLANGVLSREFEIRFQELSSVLFYSGMNTYHAYIELMPPMIYWH 235
 OY 239 ----RDGSLSDKR-----FLFTSNGCSRSLSEFPDGOIRASSSMOS 275
 DB 236 SGTRESGISIAEEEGVPLLYLVYIQKQELVODLVYATVGCSSLSLSEFPDGOIRASSSMOS 295
 OY 276 VNESGDQYHWSFGQARLDQGPMSASGDSNNHKKPREMLEIDLGKKKITGIRTTGSTOS 335
 DB 296 VNESGDQYHWSFGQARLDQGPMSASGDSNNHKKPREMLEIDLGKKKITGIRTTGSTOS 355
 OY 336 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEVFOGN 372
 DB 356 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEV-RCN 391

RESULT 7
 ID AAY79459 standard; Protein: 398 AA.
 AC AAY79459;
 DT 15-JUL-2002 (first entry)
 DE Human Neuropilin-Hy1.
 XX

KW Human: neuropilin-Hy1; chromosome 6q21; neuronal growth;
 KW nerve regeneration; neurodegenerative disease; learning disorder;
 KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
 KW organ growth; nervous system lesion; cancer; cell proliferation;
 KW cell differentiation; stem cell growth factor activity;
 KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
 KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
 KW reperfusion; food supplement; DNA microarray.
 KM
 XX
 OS Homo sapiens.
 XX
 PN WO200222780-A2.
 PD 21-MAR-2002.
 XX
 PF 11-SEP-2001; 2001WO-US285390.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-0659671.
 XX
 PA (TANG/) TANG T Y.
 PI Tang TY;
 XX
 DR WPI: 2002-351881/38.
 DR N-PSDB: ABA49565.
 XX
 PT New neuropilin-like polypeptides for diagnosing, preventing and
 PT treating neurological conditions and disorders, cancers, and for
 PT inducing angiogenesis and neovascularisation
 XX

PS Claim 3; Page 118-120; 144pp; English.
 XX
 CC The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischaemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hy1.
 XX
 SQ Sequence 398 AA:
 Query Match 37.4%; Score 1409.5; DB 23; Length 398;
 Best Local Similarity 72.8%; Pred. No. 2.6e-109;
 Matches 289; Conservative 11; Mismatches 26; Indels 71; Gaps 7;

OY 38 GDGCGHLYTYDSCGTMSTKNTPGTYPNHTVCEKTTVPKGRKRLIRLGDLDIESQTCASD 97
 DB 4 GDGCGHLYTYDSCGTMSTKNTPGTYPNHTVCEKTTVPKGRKRLIRLGDLDIESQTCASD 63
 OY 98 YLFTSSSDQGYGCGSMTPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDILT 157
 DB 64 YLFTSSSDQGYGCGSMTPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDILT 121
 OY 158 CLERSHLYKTEYSK-FCPACGRDYAGDISGMNDGYRDTSLCKRAIHAGIADDELGGQ 216
 DB 122 QGDRSEKTLDOQSRFTLATGTFVKDSFS--TDG---TSLCKRAIHAGIADDELGGQ 175
 OY 217 ISVLORKGISREGLANGVLS-----FLFTSNGCSRSLSEFPDGOIRASSSMOS 238
 DB 176 ISVLORKGISREGLANGVLSREFEIRFQELSSVLFYSGMNTYHAYIELMPPMIYWH 235
 OY 239 ----RDGSLSDKR-----FLFTSNGCSRSLSEFPDGOIRASSSMOS 275
 DB 236 SGTRESGISIAEEEGVPLLYLVYIQKQELVODLVYATVGCSSLSLSEFPDGOIRASSSMOS 295
 OY 276 VNESGDQYHWSFGQARLDQGPMSASGDSNNHKKPREMLEIDLGKKKITGIRTTGSTOS 335
 DB 296 VNESGDQYHWSFGQARLDQGPMSASGDSNNHKKPREMLEIDLGKKKITGIRTTGSTOS 355
 OY 336 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEVFOGN 372
 DB 356 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEV-RCN 391

RESULT 8
 ID AAY70539 standard; Protein: 669 AA.

XX AAY70539;
AC 04-JUL-2000 (first entry)
DT XX
XX
DE Human Factor 8 Homologue.
XX
XX Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
OS Homo sapiens.
XX
XX WO200012532-A1.
PN 09-MAR-2000.
PD
XX 20-AUG-1999; 99WO-US19047.
PF 31-AUG-1998; 98US-0098521.
PR (ELIL) LILLY & CO ELI.
XX
XX Rostec PRJ, Su W, Li XM;
PI WPI; 2000-256580/22.
DR N-PSDB; AA251872.
XX
XX Factor 8 homolog polypeptides and nucleic acids encoding them for
PT treating coagulation related disorders such as hemophilia and stroke
PS Claim 3; Page 64-66; 68pp; English.
XX
XX The present sequence is a human Factor 8 homologue (F8H),
CC a coagulation cofactor which is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
XX
SQ Sequence 669 AA;
Query Match 24.0%; Score 904; DB 21; Length 669;
Best Local Similarity 35.0%; Pred. No. 1.3e-66;
Matches 233; Conservative 109; Mismatches 217; Indels 106; Gaps 23;
OY 77 GKRLLRLADDDIE-SQTASDYLFTS-----SSDQGPCG-SMTVPKELLANTSEVY 130
DB 2 GERVRIKFEFDIEDSDCHFNFLRYNGVGRTEIGKCGGLGMNHSIEKGEIDTL 61
OY 131 RESSGSHIGRGFLTYASSDHPDLITCEBASHVLTXYEKSFCPCACRVADISGNM 190
DB 62 LFMGSHVSGRGFLASYIDKODLITCDLTASNFLEPEFSKCPMGLLPFEISGTL 121
OY 191 DGYRDTSLCKAIIHAGIADLGGQISYLRKRGISRYEGILANGVLSRDSLSDRFLE 250
DB 122 HGYRDSPLCMAGVHAGVSNLGGQISYLRKRGISRYEGILANGVLSRDSLSDRFLE 181
OY 251 TSNCGCRSLSFE-----PDQIRASSSMQSVNESGDVHNSPGARLQDQPSMASGDSN 306
DB 182 KTSGCGTIGMESGVIAIDQITASVLEWDHGOENSKPKKARLKKPPPMAPAAATDE 241
OY 307 NHRPREMLEIDGCKKTKITGRTGSGTOSNFNFYKSFVNFKNNNSKMKTYGIVNNEE 366
DB 242 ----YMWLDQIDLNKKKKITGITGSTMWENHTYVSAIRLLYSDDOCKRTVYNEPEVED 237

OY 367 KTFQNSNFRDPVQNNFPIPVARYRVNVPQTHORIALKVELLIGQ-----ITQ- 416
DB 298 KIFQGNKDYHQDVYNNFLPPIARFIVNPTOMQOKIAMKELLGCOFIPKGRPPKLTOP 357
OY 417 -----GNDSLVMKRTSGSTSVSTKKEDEFTIRPI-----PSEET--STGINITT- 458
DB 358 PPPRNSND-----LKNTPAPKIAKGRAPFTQPLQPRSSNEPPAQEQOTTASPDIRNTV 413
OY 459 -----VAIPVLVVLVYVAGMGIFAAF-----RKKRKGSS--PYGSAEAOKTDG 500
DB 414 TPNTKDVLAALAAVLVPLVAVLTLILILVCAMHNRKKKKKTEGIDLPW-----DRAG 469
OY 501 WKQIKTFE-----ARHOSAEFTISYDNEKEKTFQKLDITSMDADYQOPLMIGTG 549
DB 470 WKGMKQFLPAKAVDHEETPVYSSSE--VNHLSPREVT--TVLQADSATYQAPLVGCI 524
OY 550 TVTRKSGTFPRMPTDAEAG-VSTDAGHIDCQORAGRHEALPLAPPEEYATPVY--- 605
DB 525 GTLHORSTFRP--EEGKEAGYADLDP--YNSPGQEVYHAYABPLPTGPEVATPIIMDM 579
OY 606 ERHVLRAHTFSAOSGYRVPGPGCHKHSLSSGGFSPVAGYGAQDDGYORPHSAQPADRGY 665
DB 580 SGHPTTSVCGPSTSTFRATGNP-----PPL--VGTYNLTLSRTDSCSSAQAOT 626
OY 666 DRPKA 670
DB 627 DTPKA 631
RESULT 9
AAG75450
ID AAG75450 standard; Protein; 583 AA.
XX
XX AAG75450;
AC 03-SEP-2001 (first entry)
DT
XX
DE Human colon cancer antigen protein SEQ ID NO:6214.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
XX Homo sapiens.
OS
XX
XX WO200122920-A2.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
PF
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
DR N-PSDB; AAH34855.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7657-7660; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions

In a patient's genome that affect the activity of p by expressing inactive proteins or to supplement the patients own production of p. Additionally, N may be used to produce the colon cancer-associated ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 CC and AAB7789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 662 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1021 to 1052, 7921 and 7922.

	SQ	Sequence	583 AA;
Query Match	Best Local Similarity	23.9%; Score 902; DB 22:	Length 583;
Matches	Conservative	85; Mismatches	170; Indels
Gaps	13		
Dy	20	LALLAASAPRLQAEBLDCGGLVYDQGSMATSKNYPGTYPNHTVCERTITVRGRK R 79 : : : : : : : :	
Dz	29	LLLLLVLTLLEDDAAOOGCGGHLYLGESGTLSINYPQTVPNSTVEWEIRVMGER B 88 	
Oy	80	LILRIGDLIE-SQCASDYLFETSS-QSYGYPC-SMTVPRELLINTSEVAREE I 133 : : : : : : : SSOXYGPC-SMTVPRELLINTSEVAREE I 133 : : : : : : : SSOXYGPC-SMTVPRELLINTSEVAREE I 133	
Dd	89	VRIKGDPEDIESDSCHFNFLRYIVNGIGVSFEIYGCCGLQMNSISKNGNETILLFM 148 	
Oy	134	SGSHFSGRGFTLVYASDSPDLITCLERASHLKTEYSKEFCPACGDVAAGDISNMVDGY 193 : : :~::~:: : :~::~:: : :~::~:: : :~::~: 149 SGIHVSGGFSLASYEVTDKDQLTICHDNANLNLEPFERSYCPCAGCPLPAELISGITPHCY 208 	
Oy	194	RDTSLCKRAAHAGIADELGCQISYLARKKSIRREGIIANGVLSRDGSLSDKRFLTSTN 253 : ~ :: : : : : : : : : : : 209 RDSSLCPAMAGAVHASNTLGGQISVVISKGIPIYEISSLANNVTSVGHLSTGLEFKTS 268 : ~ :: : : : : : : : : : :	
Oy	254	GCSRSLSE---PGQAIRASSMSQSVNESGDQVHMSPGOARLDOPGPMAAGDSSNNHK 309 : : : : : : : : : : : 269 GCCTGLMGESGVIANPDITASLTMDTHTGDNEMSKPKAKALKRKRPWMAAFDE---- 325 	
Oy	310	PHEWLIEDGEKKKITGTINTGSTOSNFNEFYKSFVNENKNNSKMKTYYKGIYNNEKYF 369 : : : : 326 -YOMLOIDLNRKKTKTGTITTGTIEWEHYVASAYRILTSDDGOKVTWRREGVEDDKIF 384 	
Oy	370	OONSFPPOPVONFFPTIPYRVVRVPPWHORIALVKYLICO-----ITO----- 416 : : 385 OGKNDYHODVRRKAFLPRIIAKTRRNPRPMOOCKIAMKMLLCQRTLPKRAPRKLTOPPPP 444 	
Oy	417	-GNDSLWKRTSOSTSVSTRKEDETFRPI-----PSET-STGINITY- --- 458 : : : : : : : : : : : 445 RMSND---LNKTPATPKARIAKRAFEPQPDRPSNSEPEPATGETASPDIRNTVTBN 500 	
Oy	459	--VAIFLVYLTVVFAGMGIFAF-----RKKKKGS--PY 490 : : : : : : : : : : : 501 VTKDVLAALVALVPLVLMVLTTLILLYCAMHRNRRKKTKEYIDLPY 547 	
RESULT 10	AEE22716	AAE22716 standard; Protein; 385 AA.	
ID	AAE22716	standard; Protein; 385 AA.	
XX	AAE22716;		
XX	09-AUG-2002	(first entry)	
XX			
DE		Human neuropilin-Hy2 protein.	
KW		Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation; neurodegenerative disease; Alzheimer's disease; learning; angiogenesis; chromocytoma; memory; platelet; plastic anaemia; antiinflammatory; ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoicide; wound healing; tissue repair; Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; Shy-Draeger syndrome; trauma; spinal cord; cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis; human immunodeficiency virus; HIV:autoimmune disorder; dermatological;	

KW		systemic lupus erythematosus; rheumatoid arthritis; antihypertoid; cancer;
KV		autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KM		myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KL		neurotropic; neuroprotective; vulvovaginal; anticoagulant; antiparasitic;
KJ		cerebroprotective; tranquilliser; virocidal; antibacterial; cytostatic;
KI		immunosuppressive; chromosome 6q21.
XX		
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	Peptide	1..20
FT	/label= Signal_peptide	
FT	Protein	61..385
XX		/note= "Human mature neuropilin-Hy2 protein"
PX	WO200222815-A1.	
PD		
XX	21-MAR-2002.	
XX		
XX	12-SEP-2001; 2001WO-US28488.	
PR	11-SEP-2000; 2000US-0659671.	
PR	06-SEP-2001; 2001US-317902P.	
XX		
PA	(HXSE-) HXSEQ INC.	
PI	Tang YT;	
DR	WIPI: 2002-393966/42.	
DR	N-PDB; AAD3594.	
PT	Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides	
PT	useful for treating neurodegenerative diseases e.g. Alzheimer's	
PT	disease, and for diagnosing and mapping genetic neuronal defects	-
XX		
PS	Claim 3; Page 128-130; 152pp; English.	
XX		
XX	The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and	
CC	neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-	
CC	like polypeptides and polynucleotides are useful in modulating neuronal	
CC	growth regenerative capacity, treating neurodegenerative diseases,	
CC	diagnosing and mapping genetic neuronal defects and degenerative diseases	
CC	like Alzheimer's disease and for treating learning and memory disorders.	
CC	They are also useful for inducing angiogenesis, neovasculatisation, as	
CC	well as organ growth and development e.g. heart and other tissues.	
CC	Antagonists of neuropilin-like polypeptides are useful for treating	
CC	cancers and other malignant diseases. Neuropilin is used to treat	
CC	platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal	
CC	nocturnal haemoglobinuria and is used in nerve tissue growth or	
CC	regeneration, in wound healing, tissue repair and replacement and in	
CC	healing of bones, incisions and ulcers. Compositions comprising the	
CC	sequences of the invention are useful for treating diseases of peripheral	
CC	nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager	
CC	syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.	
CC	spinal cord disorders, head trauma and cerebrovascular diseases e.g.	
CC	stroke, ulcers, immune deficiencies and immune disorders, infections by	
CC	hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,	
CC	mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.	
CC	multiple sclerosis, systemic lupus erythematosus, Rheumatoid arthritis,	
CC	autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes	
CC	mellitus, graft-versus-host disease, myasthenia gravis and autoimmune	
CC	inflammatory eye disease. The nucleic acids of the invention are used in	
CC	gene therapy techniques. The present sequence is human neuropilin-Hy2	
CC	protein. Neuropilin-Hy2 gene is located on chromosome 6q21.	
XX		
SQ	Sequence	385 AA;
XX		
Query Match	18.4%; Score 692.5; DB 23; Length 385;	
Best Local Similarity	41.8%; Pred. No. 3e-49;	
Matches	163; Conservative 65; Mismatches 131; Indels 31; Gaps 13	
20	LALLVASPLRLQAEEIDGCGHLVTYDSGTMTSKNPGTYPNHTVCERTITVPKGR	79
I	: :	1-1

Db 5 LLLLVLLLEEDAGAGQGGCGHTVAGPESGLTINIPQIYPNSTVCMEIRVKMER 64
 QY 80 LILRLGDLIE-SQTCASDYLFTS-----SSDQYGPCG-SMTVPKELLINTSEVTYRFE 133
 Db 65 VRIKFGDFIEDSDSCHFNRLRYNGIGVSRTEIGKXCGIGLOMNSISEKGEITLTFM 124
 QY 134 SGGHISGRGFLTYASSDHPDLITCLERASHYKTEYSKFCPCAGCDVADISGNMVDGY 193
 Db 125 SGHVSGRGLFASYSYIDKODLITCLDTASNFLEPESKCPAGCLLPFAEISGTFPHGY 184
 QY 194 RDTSLCKAAIHAGIADLEGQISVLRKGISRYEGIIANGVLSRDSLSDRFLETSN 253
 Db 185 RDSSPLCMAGVHAGVSNLTGGQISVYISKIGIPYESSLANNTSVYGHLSLTFPKTS 244
 QY 254 GCSRSLSFEPDQGIIRASS-----WQSVNESGDVHNSPGQARLQDQPSW-ASGDS 305
 Db 245 GCYGTIGMESGPD-RGSSNNSITVLEW--TDHNGOESKMPKKSQAE--KTYTALGAF 298
 QY 306 NNHKPREMLEIDGEEKKKTIGRTGSTQ-SNENFYKVS--FVNNFKNNNSKWKTYKGI 361
 Db 299 TDE--YQWLOIDLNKKKKTIGITGSTWVSTITMCLPTESCIVMMGRNG----LCESL 352
 QY 362 VNNEEKVFOGNSN-FRDPVONNFIPIVAR 390
 Db 353 VVEQDKIFQGNKRIITRMVANNFLPIIAR 382
 RESULT 11
 AAU79460
 ID AAU79460 standard; Protein; 385 AA.
 XX
 AC AAU79460;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human Neuropilin-Hy2.
 XX
 KW Human; neuropilin-Hy2; chromosome 6q21; neuronal growth;
 KW nerve regeneration; neurodegenerative disease; learning disorder;
 KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
 KW organ growth; nervous system lesion; cancer; cell proliferation;
 KW cell differentiation; stem cell growth factor activity;
 KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
 KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
 KW reperfusion; food supplement; DNA microarray.
 XX
 OS Homo sapiens.
 FH Key
 FT Peptide 1..20 Location/Qualifiers
 FT Protein 21..385 /label= Signal_peptide
 FT /label= Mature_neuropilin_Hy2
 XX
 PN WO200222780-A2.
 PD 21-MAR-2002.
 XX
 PF 11-SEP-2001; 2001WO-US28590.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-0659671.
 XX
 PA (TANG/) TANG T Y.
 XX
 PI Tang TY;
 XX
 DR WPI; 2002-351881/38.
 DR N-PSDB; ABR49567.
 XX
 PT New neuropilin-like polypeptides for diagnosing, preventing and
 treating neurological conditions and disorders, cancers, and for

PT Inducing angiogenesis and neovascularisation -
 XX
 XX Claim 3; Page 123-125; 144p; English.

CC The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischaemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hy2.
 XX
 XX Sequence 385 AA;

Query Match 18.4%; Score 692.5; DB 23; Length 385;

Best Local Similarity 41.8%; Pred. No. 3e-49;

Matches 163; Conservativity 65; Mismatches 131; Indels 31; Gaps 13;

QY 20 LALLVAVSAPRLVQAEELGDCGHLTYVODSGTMSKNPYGTYPNHTVCERTYVPRGKR 79
 Db 5 LLLLVLLLEEDAGAGQGGCGHTVAGPESGLTINIPQIYPNSTVCMEIRVKMER 64
 QY 80 LILRLGDLIE-SQTCASDYLFTS-----SSDQYGPCG-SMTVPKELLINTSEVTYRFE 133
 Db 65 VRIKFGDFIEDSDSCHFNRLRYNGIGVSRTEIGKXCGIGLOMNSISEKGEITLTFM 124
 QY 134 SGGHISGRGFLTYASSDHPDLITCLERASHYKTEYSKFCPCAGCDVADISGNMVDGY 193
 Db 125 SGHVSGRGLFASYSYIDKODLITCLDTASNFLEPESKCPAGCLLPFAEISGTFPHGY 184
 QY 194 RDTSLCKAAIHAGIADLEGQISVLRKGISRYEGIIANGVLSRDSLSDRFLETSN 253
 Db 185 RDSSPLCMAGVHAGVSNLTGGQISVYISKIGIPYESSLANNTSVYGHLSLTFPKTS 244
 QY 254 GCSRSLSFEPDQGIIRASS-----WQSVNESGDVHNSPGQARLQDQPSW-ASGDS 305
 Db 245 GCYGTIGMESGPD-RGSSNNSITVLEW--TDHNGOESKMPKKSQAE--KTYTALGAF 298
 QY 306 NNHKPREMLEIDGEEKKKTIGRTGSTQ-SNENFYKVS--FVNNFKNNNSKWKTYKGI 361
 Db 299 TDE--YQWLOIDLNKKKKTIGITGSTWVSTITMCLPTESCIVMMGRNG----LCESL 352
 QY 362 VNNEEKVFOGNSN-FRDPVONNFIPIVAR 390
 Db 353 VVEQDKIFQGNKRIITRMVANNFLPIIAR 382

RESULT 12

ID AAW96308 standard; Protein: 925 AA.
AC AAW96308;
XX 28-JUN-1999 (first entry)
XX Neuropllin-2.
XX Neuropllin; neuropilin-2; semaphorin; transmembrane protein; axon;
XX neurone; development; cell growth; immune response;
XX viral pathogenesis; treatment; disease; graft rejection;
XX viral disease; oncological disease; cancer; screening; probes.
XX Rattus rattus.
OS
XX MO9904263-A1.
XX 28-JAN-1999.
XX 17-JUL-1998; 98WO-US14632.
XX 17-JUL-1997; 97US-0052762.
XX (UyJo) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX Glincy DD, Kolodkin AL;
XX WPI; 1999-132446/1.
XX N-PSDB; AAX08417.
XX New isolated semaphorin receptor, neuropilin-2 - used to develop
XX products for the diagnosis and treatment of neurological,
XX immunological, oncological and viral diseases
XX
XX Claim 13; Page 77-79; 88pp; English.
XX
XX The neuropilins are type I transmembrane proteins and act as
XX semaphorin III (Sema III) receptors. The semaphorins have been
XX shown to function in repulsive axon guidance. Sema III is a
XX secreted protein that in vitro causes neuronal growth cone
XX collapse and chemorepulsion on neurites and is required in vivo for
XX correct sensory afferent innervation and other aspects of
XX development. Agents which inhibit or enhance the interaction of a
XX semaphorin and a neuropilin can be potent modulators of nerve cell
XX growth, immune responsiveness, and viral pathogenesis, and can be
XX used in the treatment and diagnosis of neurological disease,
XX neuro-regeneration, immune modulation including hypersensitivity and
XX graft-rejection, and diagnosis and treatment of viral and oncological
XX infection/diseases. The neuropilins, neuropilin-encoding nucleic
XX acids, and unique portions also are useful in screening chemical
XX libraries for regulators of semaphorin-mediated cell activity, and in
XX genetic mapping as probes for related genes, as diagnostic reagents
XX for genetic, neurological, immunological and oncological disease.
XX
SQ Sequence 925 AA:
Query Match 9.6%; Score 362.5; DB 20; Length 925;
Best local Similarity 22.3%; Pred. No. 5.2e-21;
Matches 143; Conservative 100; Mismatches 246; Indels 151; Gaps 24;
OY 41 CGHLYTVDDSTMTSKNPGTYPNITVEKITIYVK-GKRLILRLG-DLDESQCASDY 98
DB 28 CGGRILNSKDAQYITSPGYDPSHQCEWVYVAEPNOKIVLNPNPHEIEKHCKKYDF 87
OY 99 ILF-----TSSSDQYGPYCGSMTVPKELLANTSEVTRFESGSHISGRGFLTY-----AS 149
DB 88 IEIRDGSESEADLAKHCGNT-APPITIISSGSVLYIKFTSDYAROGAGFSLRYELFKTGS 146
OY 150 SD-----HPDLITC-----IERASHYLKTEYSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFPEKPYPHNIDCTFTIILAKPRMEIILQFLTFDLHDPLQVCEG 206
OY 174 -CPAGCRVADIS--GMMVDGYRDTSLCKRAIHAIIADELGQIVYLQRKGISRY-- 228

DB 207 DCKYMDLIMDGIPIHVGLPIKCYCGTKPPSKIRSRSTGILSLTFHTDMAVADGFSARYLL 266
OY 229 -----EGILANGVYSRD-GSLSDKRFLFTSGSGRSLSFEPDGOIRASSSQSVNESGD 281
DB 267 VHQEPPENFQCNAPLGMESGRIANE-----QISASSTF-----S 300
OY 282 QVHWSPGQARLQDQGPSWASGDSNNHPRMELIDGKKKITGIRTTG--STOSNENF 339
DB 301 DGRMTPOQSRLHGDNDNGWTPNVDSN---KEYLQVDLFLFMLPAIATGALISRETQCY 356
OY 340 YKSPVAMFKNNKSKWKYKIGLVNNEEYVFOGNSFRPQONNFPPIVAVYRVPTW 399
DB 357 YKSYKLEVSTNGEDMAYVRRGRNH--RVQANNDATELVNKLHTPLTRIRIPQTW 414
OY 400 HORIALKVELIGCOITQGNDSLVMKRTS---QSFVSSTKKDEITTRPI----- 445
DB 415 HIGIALRLELEGCRTDAPCSNMLGSLADTQISASTREIYMSAARLYVSRSGW 474
OY 446 -----PSEETSTGINTTVAIPLVLLVYLVFAGMGIFA---AFRKKRKKSGPYGA 493
DB 475 FPRNPQAPGGEW-LQVDLGPPTVKYGVIIQARGGDSITAMEARAFAVKKFKVYSLNGK 533
OY 494 EAQKTDCKKQIKYPPARQSA-EFTISYD-----NEKEMTOKL 530
DB 534 D-----WEYIDPPTQPKLEEGNMHYDTPDIRFEFVPAQYRVYPERMSPAGIGRL 587
OY 531 DLITSDMADYQOPL-MIGTGVTRKSGTFRPMDDTDAEEAG 569
DB 588 EVLGCDWTDSKPVTVELGPTVKSSEETTPPYPMDEBATECG 627

Search completed: May 15, 2003, 13:20:42
Job time : 68.5731 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:15:30 ; Search time 24.0681 Seconds

(without alignments)
2865.294 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MWPARGGALARAAGRL.....YSAPDCLPLNQTAMTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 362588 segs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications.AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3770	100.0	715	US-10-003-132-2	Sequence 2, Appl1
2	3770	100.0	715	US-09-759-130B-73	Sequence 73, Appl1
3	3610	95.8	681	US-09-759-130B-75	Sequence 75, Appl1
4	2229	59.1	421	US-09-759-130B-76	Sequence 76, Appl1
5	1924.5	51.0	503	US-10-003-132-4	Sequence 4, Appl1
6	1912.5	50.7	503	US-09-823-038A-51	Sequence 51, Appl1
7	1650	44.8	458	US-10-003-132-6	Sequence 6, Appl1
8	1265	33.6	729	US-09-759-130B-78	Sequence 78, Appl1
9	1042	27.6	725	US-10-060-830-3	Sequence 3, Appl1
10	873.5	23.2	653	US-10-060-830-1114	Sequence 1114, Appl1
11	351.5	9.3	931	US-10-104-440-4	Sequence 4, Appl1
12	335	8.9	923	US-10-104-440-2	Sequence 2, Appl1
13	317	8.4	2224	US-10-115-563-14	Sequence 14, Appl1
14	306.5	8.1	1431	US-10-095-718-4	Sequence 4, Appl1
15	299	7.9	343	US-10-190-593-4	Sequence 2, Appl1
16	299	7.9	387	US-10-190-593-4	Sequence 4, Appl1
17	285.5	7.6	1438	US-10-225-900-1	Sequence 1, Appl1
18	285.5	7.6	1438	US-10-006-091-1	Sequence 1, Appl1
19	285.5	7.6	1438	US-10-047-257-1	Sequence 1, Appl1

20	285.5	7.6	1471	12	US-10-095-718-2	Sequence 2, Appl1
21	285.5	7.6	2332	9	US-09-957-641-2	Sequence 2, Appl1
22	285.5	7.6	2332	9	US-10-187-319-2	Sequence 2, Appl1
23	285.5	7.6	2332	9	US-10-132-829-4	Sequence 4, Appl1
24	276.5	7.3	2319	9	US-10-187-319-6	GENERAL INFORNA
25	271.5	7.2	1443	9	US-10-187-319-39	Sequence 39, Appl1
26	271.5	7.2	2133	9	US-10-187-319-37	Sequence 37, Appl1
27	226	6.0	1128	9	US-09-996-015-10	Sequence 10, Appl1
28	226	6.0	1128	9	US-10-238-876-8	Sequence 8, Appl1
29	217	5.8	845	9	US-09-996-015-11	Sequence 11, Appl1
30	217	5.8	1159	9	US-10-238-876-2	Sequence 2, Appl1
31	216.5	5.7	718	9	US-09-996-015-9	Sequence 9, Appl1
32	185.5	4.9	250	10	US-09-770-643A-8	Sequence 8, Appl1
33	185.5	4.9	279	10	US-09-770-643A-10	Sequence 10, Appl1
34	185.5	4.9	534	10	US-09-770-643A-14	Sequence 14, Appl1
35	185.5	4.9	582	10	US-09-770-643A-12	Sequence 12, Appl1
36	185.5	4.9	697	10	US-09-770-643A-18	Sequence 18, Appl1
37	185.5	4.9	745	10	US-09-770-643A-16	Sequence 16, Appl1
38	185.5	4.9	791	10	US-09-770-643A-20	Sequence 20, Appl1
39	185.5	4.9	839	10	US-09-770-643A-22	Sequence 22, Appl1
40	185.5	4.9	1259	10	US-09-770-643A-4	Sequence 4, Appl1
41	185.5	4.9	1307	10	US-09-770-643A-2	Sequence 2, Appl1
42	184.5	4.9	1298	10	US-09-770-643A-24	Sequence 24, Appl1
43	184.5	4.9	1298	10	US-09-770-643A-28	Sequence 28, Appl1
44	182.5	4.8	764	10	US-09-900-518A-2	Sequence 2, Appl1
45	180.5	4.8	686	10	US-09-874-198-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Shoenker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

Query Match 100.0%; Score 3770; DB 9; Length 715;
Best Local Similarity 100.0%; Pred. No. 8e-284;
Matches 715; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWPARGGALARAAGRLALILAVASAPRLAEELEGDCGHLVYQDSGTMTSKNYPG 60
DB 1 MWPARGGALARAAGRLALILAVASAPRLAEELEGDCGHLVYQDSGTMTSKNYPG 60
QY 61 TYPNHTVCERTITVPKGRILRLRGDIDISQTCASDYLLFTSSSDQYGPYCSMTYPK 120
DB 61 TYPNHTVCERTITVPKGRILRLRGDIDISQTCASDYLLFTSSSDQYGPYCSMTYPK 120
QY 121 LLNTSEVTVRFESGSHISGRGLTYASSDHPDLITCLERASHYLTETYSKPCPACRD 180
DB 121 LLNTSEVTVRFESGSHISGRGLTYASSDHPDLITCLERASHYLTETYSKPCPACRD 180
QY 181 VAGDISNMVDGYRDPISLCKAIIHAGITIDEIGGOISYORGISYBEIILANGVLSRD 240
DB 181 VAGDISNMVDGYRDPISLCKAIIHAGITIDEIGGOISYORGISYBEIILANGVLSRD 240

```

OY 241 GSLSDKRLFTSNGCSRSLSEFPDQOIRASSSMOSVNSGQVHMSPGQARLDOQGPSMA 300
D 241 GSLSDKRLFTSNGCSRSLSEFPDQOIRASSSMOSVNSGQVHMSPGQARLDOQGPSMA 300
OY 301 SGDSNNHKKPREMLEIDGEEKKITGITRTGSTOSNENFYKSFVNMKNNSKMKTKG 360
D 301 SGDSNNHKKPREMLEIDGEEKKITGITRTGSTOSNENFYKSFVNMKNNSKMKTKG 360
OY 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPOTHORIALKVELICQITQGNDS 420
D 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPOTHORIALKVELICQITQGNDS 420
OY 421 LVMRTSGSTSVSTRKKEDETTTRPIPSEETSTGINITVAIPVLVLYVFRAGMIFAF 480
D 421 LVMRTSGSTSVSTRKKEDETTTRPIPSEETSTGINITVAIPVLVLYVFRAGMIFAF 480
OY 481 RKKKKGSPYSAQAOKTDCWKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
D 481 RKKKKGSPYSAQAOKTDCWKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
OY 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHIDCQORAGRHEYALPLAPEPEY 600
D 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHIDCQORAGRHEYALPLAPEPEY 600
OY 601 ATPYERHVLAAHTFSAGSYRVPGOPGKHSLSSGGFSPVAGVAGDGDYORPHSQP 660
D 601 ATPYERHVLAAHTFSAGSYRVPGOPGKHSLSSGGFSPVAGVAGDGDYORPHSQP 660
OY 661 ADRCYDRKAVSALATESGHPDSOKPPTHPGTSYSAPRDLPLNOTANTALL 715
D 661 ADRCYDRKAVSALATESGHPDSOKPPTHPGTSYSAPRDLPLNOTANTALL 715

RESULT 2
US-09-759-130B-73
: Sequence 73, Application US/09759130B
: Publication No. US2003022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D
: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirst, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wighton, Nicolas
: APPLICANT: Goodearl, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OR INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OR INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OR INVENTION: USES.
: FILE REFERENCE: MP100-5350NIM
: CURRENT APPLICATION NUMBER: US/09/759, 130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479, 249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559, 497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578, 063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333, 159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596, 194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342, 364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608, 452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393, 996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602, 871

```

```

: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420, 707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 73
: LENGTH: 715
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-759-130B-73

Query Match      100.0%; Score 3770; DB 9; Length 715;
Best Local Similarity 100.0%; Pred. No. 8e-284;
Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYPGARGGALARAAGRLALLAVASAPLRIQAEELEDGCGHLYTQDSTMTSKNPG 60
D 1 MYPGARGGALARAAGRLALLAVASAPLRIQAEELEDGCGHLYTQDSTMTSKNPG 60
OY 61 TYPNHTVEKTIITYPKGRLLRLRGDDIEDSQCASDYLFTSSSDOYGYCGSKNYPKE 120
D 61 TYPNHTVEKTIITYPKGRLLRLRGDDIEDSQCASDYLFTSSSDOYGYCGSKNYPKE 120
OY 121 LLLNTSEVTVFEESGSHISGRGFLITVASSDPDLITCERASHYLTKEYSKFCPAGCRD 180
D 121 LLLNTSEVTVFEESGSHISGRGFLITVASSDPDLITCERASHYLTKEYSKFCPAGCRD 180
OY 181 VAGDISGMVNDGYNDTSLCKAAIHAGIADDELGGQISVLRKGISIREGILANGVLSRD 240
D 181 VAGDISGMVNDGYNDTSLCKAAIHAGIADDELGGQISVLRKGISIREGILANGVLSRD 240
OY 241 GSLSDKRLFTSNGCSRSLSEFPDQOIRASSSMOSVNSGQVHMSPGQARLDOQGPSMA 300
D 241 GSLSDKRLFTSNGCSRSLSEFPDQOIRASSSMOSVNSGQVHMSPGQARLDOQGPSMA 300
OY 301 SGDSNNHKKPREMLEIDGEEKKITGITRTGSTOSNENFYKSFVNMKNNSKMKTKG 360
D 301 SGDSNNHKKPREMLEIDGEEKKITGITRTGSTOSNENFYKSFVNMKNNSKMKTKG 360
OY 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPOTHORIALKVELICQITQGNDS 420
D 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPOTHORIALKVELICQITQGNDS 420
OY 421 LVMRTSGSTSVSTRKKEDETTTRPIPSEETSTGINITVAIPVLVLYVFRAGMIFAF 480
D 421 LVMRTSGSTSVSTRKKEDETTTRPIPSEETSTGINITVAIPVLVLYVFRAGMIFAF 480
OY 481 RKKKKGSPYSAQAOKTDCWKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
D 481 RKKKKGSPYSAQAOKTDCWKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
OY 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHIDCQORAGRHEYALPLAPEPEY 600
D 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHIDCQORAGRHEYALPLAPEPEY 600
OY 601 ATPYERHVLAAHTFSAGSYRVPGOPGKHSLSSGGFSPVAGVAGDGDYORPHSQP 660
D 601 ATPYERHVLAAHTFSAGSYRVPGOPGKHSLSSGGFSPVAGVAGDGDYORPHSQP 660
OY 661 ADRCYDRKAVSALATESGHPDSOKPPTHPGTSYSAPRDLPLNOTANTALL 715
D 661 ADRCYDRKAVSALATESGHPDSOKPPTHPGTSYSAPRDLPLNOTANTALL 715

RESULT 3
US-09-759-130B-75
: Sequence 75, Application US/09759130B
: Publication No. US2003022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D

```

```

: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirt, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodheart, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OF INVENTION: USES
: FILE REFERENCE: MP100-5350NMIM
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 75
: LENGTH: 681
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-759-130B-75

Query Match      95.8%; Score 3610; DB 9; Length 681;
Best Local Similarity 100.0%; Pred. No. 1.9e-271;
Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirt, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodheart, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OF INVENTION: USES
: FILE REFERENCE: MP100-5350NMIM
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 76
: LENGTH: 421
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-759-130B-76

RESULT 4
US-09-759-130B-76
: Sequence 76, Application US/09759130B
: Publication No. US20030022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D
: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirt, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodheart, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OF INVENTION: USES
: FILE REFERENCE: MP100-5350NMIM
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 76
: LENGTH: 421
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-759-130B-76

```

```

Db      125  SGNHSISKGFLLTYAASSDHPDLITCLENGSHYPEEKYSKPCFAGCRDILAGDISGNTKGY 184
Qy      194  RDTSLTCKAAIHAGIITADELGGQISVLOKRGISRTREGILANGVLSRSGSLSDKRFLETSN 253
Db      185  RDTSLTCKAAIHAGIITADELGGHINLQSKGISHEGILANGVLSRHSLSEKRFLEF--- 241
Qy      254  GCSRLSPEDPGQIRASSSWGVSNEGQVHMSPOQARLQOQGPGWAGSDSNHKKPREW 313
Db      242  ----- 241
Qy      314  LEIDLGEKKITGITRTGSTOSNFYKSFVNFKNNSKWKTYKGIYNNEKVFQNS 373
Db      242  ----- 241
Qy      374  NERDPVONNEIPPIVARIYRVVPOTWIORIALKVELLIGQITQGNDSLWKRTSOSTSVS 433
Db      242  ----- 241
Qy      434  TKKDEITTRIPSEDETSGITNTVAILPLVLYLVFAGMGIFAFAKRRKKKSPYSA 493
Db      242  -----TTPGKNITTVALPSVIFIALLLTGMGITAIKRRKKKGPYSA 285
Qy      494  EAOQTDCKWQKIQIYPFARHQAFTISYDNEKEMOKLDLITSDMADYQOPLMIGITVR 553
Db      286  DAQGTGCMKQKQIYFARHQAFTISYDNEKEMOKLDLITSDMADYQOPLMIGITVR 345
Qy      554  KGSFTFPMQDIAEEAGVSTDAAGHYDCPQARARHEYALPLARPEEYATPIVERHVLRAH 613
Db      346  KGSFTFPMQDTEEVRYVTEASGHYDCPQARARHEYALPLARPEEYATPIVERHVLRAH 405
Qy      614  TFSQSGYRVGQPGQPKHSHSSGGFSVAVGVAGQDGYORPHSAQPRDKRYDRPKAUSA 673
Db      406  TFSQSGYRVGQPRPKHSHSSGGFSVAVGVAGQDGYORPHSAQPRDKRYDRPKAUSA 464
Qy      674  LATESGHPSQKPEPTHPTSDSYGAPRDCPLPLNOTATALL 715
Db      465  L--DSRDPASQSQMT--SGGDDGYAPRNGIAPLPLNOTATALL 503

RESULT 6
US-09-823-038A-51
| Sequence 51, Application US/09823038A
| Patient No. US20020058335A1
| GENERAL INFORMATION:
| APPLICANT: Strachan, Lorna
| APPLICANT: Sleeman, Matthew
| APPLICANT: Abernethy, Nevlin
| APPLICANT: Onrust, Rene
| APPLICANT: Kumble, Anand
| APPLICANT: Morrison, Greg
| TITLE OF INVENTION: Compositions isolated from Stromal Cells
| FILE REFERENCE: 11000.1037c3
| CURRENT APPLICATION NUMBER: US/09/823.038A
| CURRENT FILING DATE: 2001-07-09
| NUMBER OF SEQ ID NOS: 61
| SOFTWARE: FASTSEQ for Windows Version 4.0
| SEQ ID NO 51
| LENGTH: 503
| TYPE: PRF
| ORGANISM: Mouse
US-09-823-038A-51

Query Match          50.7%; Score 1912.5; DB 10; Length 503;
Best Local Similarity 55.6%; Pred. No. 5e-140;
Matches 390; Conservative 35; Mismatches 74; Indels 203; Gaps 4;

Qy      14  AAGGALLALLAASAPRLQAEEELGDCGHLVYQDSGTMSTSKNYPGTYPHNHYCEKTTT 73
Db      5  AGGGSVALLEAVCAPIRLQAEEELGDCGHLVTSQDSGTMSTSKNYPGTYPHNHYCEKTTT 64
Qy      74  VPKRKLRLRLGDLDISQTCASDYLLFTSSSDOYGPYCGSMYPKELLMTSEYTVAFPE 133
Db      74  VPKRKLRLRLGDLDISQTCASDYLLFTSSSDOYGPYCGSMYPKELLMTSEYTVAFPE 133

```


Db 65 VPKGRLLRLGDLNLSKTCASDYLLFPSSATQOYGCSSMAVPKRLNLSNEVTYLFK 124
 QY 134 SGRHSRGLLTYYASSDHDLLITCLERASHYLTETSEKFCPCAGCRVADISGNMVDY 193
 Db 125 SGRHSRGLLTYYASSDHDLLITCLERASHYLTETSEKFCPCAGCRVADISGNMVDY 184
 QY 194 RDTSLCKAAIHAGIIDELEGQISVLRKRGISREYGIILANGVLSRDSLSDRKFLTSN 253
 Db 185 RDTSLCKAAIHAGIIDELEGQISVLRKRGISREYGIILANGVLSRDSLSDRKFLTSN 241
 QY 254 GCSRSLSFEEDGOIRASSSQSVNESGDQYHWSPGQARLQDQGPMSAGSSNNHKKREW 313
 Db 242 ----- 241
 QY 314 LEIDLGEKKKITGIRTTGTSQSNFNFYKSFVNFKNNSKMTYKGIYNNKEKVFQGS 373
 Db 242 ----- 241
 QY 374 NFRDPVQNNFIPPIVARYVRVPQTHQRIALAVELIGQITQGNDSLVMKRTSQSTSVS 433
 Db 242 ----- 241
 QY 434 TKREDETTPRIPSEETSTGINTTVAIPLVLVLVFAGMGIFAAFRKKKKGSPYGA 493
 Db 242 ----- 285
 QY 494 EAQKTDCKWKQIKYPPFARHQAETIISYDNKEKMTOKLDLITSDMADYQOPLMIGTGVTR 553
 Db 286 DAQKTDCKWKQIKYPPFARHQAETIISYDNKEKMTOKLDLITSDMADYQOPLMIGTGVTR 345
 QY 554 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHVALPLAPEPEXATPIVERHVLRAH 613
 Db 346 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHVALPLAPEPEXATPIVERHVLRAH 405
 QY 614 TFSAGSYRVGPQPGHKHSLSSGGSFPVAGVAGQDQYRPHSAQPADRGYDRPRAVSA 673
 Db 406 TFSAGSYRVGPQPGHKHSLSSGGSFPVAGVAGQDQYRPHSAQPADRGYDRPRAVSA 464
 QY 674 LATESGHPDSQKPTHTGTSDSYAPRDCITPLNQMTALL 715
 Db 465 L-DSRDPASQSQMT-SGDDGYSAPRNGLAPLNQMTALL 503
 RESULT 7
 US-10-003-132-6
 ; Sequence 6, Application US/10003132
 ; Publication No. US20020192750A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Gao, Zeren
 ; APPLICANT: Shoemaker, Kimberly E.
 ; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
 ; FILE REFERENCE: 00-62
 ; CURRENT APPLICATION NUMBER: US/10/003,132
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/249,004
 ; PRIOR FILING DATE: 2000-11-15
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 458
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-003-132-6

Query Match 44.8%; Score 1690; DB 9; Length 458;

Best Local Similarity 50.4%; Pred. No. 7, 8e-123; Mismatches 69; Indels 248; Gaps 5;

QY 14 AAGRGTLALLAYSAPIRLQAEELGDCGHLVYODSGTWSKNYPGTYPNHTVCERTIT 73
 Db 5 AAGPSVALLAFVACAPLRLQAEELGDCGHLVYODSGTWSKNYPGTYPNHTVCERTIT 64

QY 74 VPKGRLLRLGDLNLSKTCASDYLLFTSSDQOYGCSSMTVPKRLNLSNEVTYLFK 133
 Db 65 VPKGRLLRLGDLNLSKTCASDYLLFTSSDQOYGCSSMTVPKRLNLSNEVTYLFK 99
 QY 134 SGRHSRGLLTYYASSDHDLLITCLERASHYLTETSEKFCPCAGCRVADISGNMVDY 193
 Db 100 ----- 139
 QY 194 RDTSLCKAAIHAGIIDELEGQISVLRKRGISREYGIILANGVLSRDSLSDRKFLTSN 253
 Db 140 RDTSLCKAAIHAGIIDELEGQISVLRKRGISREYGIILANGVLSRDSLSDRKFLTSN 196
 QY 254 GCSRSLSFEEDGOIRASSSQSVNESGDQYHWSPGQARLQDQGPMSAGSSNNHKKREW 313
 Db 197 ----- 196
 QY 314 LEIDLGEKKKITGIRTTGTSQSNFNFYKSFVNFKNNSKMTYKGIYNNKEKVFQGS 373
 Db 197 ----- 196
 QY 374 NFRDPVQNNFIPPIVARYVRVPQTHQRIALAVELIGQITQGNDSLVMKRTSQSTSVS 433
 Db 197 ----- 196
 QY 434 TKREDETTPRIPSEETSTGINTTVAIPLVLVLVFAGMGIFAAFRKKKKGSPYGA 493
 Db 197 ----- 240
 QY 494 EAQKTDCKWKQIKYPPFARHQAETIISYDNKEKMTOKLDLITSDMADYQOPLMIGTGVTR 553
 Db 241 DAQKTDCKWKQIKYPPFARHQAETIISYDNKEKMTOKLDLITSDMADYQOPLMIGTGVTR 300
 QY 554 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHVALPLAPEPEXATPIVERHVLRAH 613
 Db 301 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHVALPLAPEPEXATPIVERHVLRAH 360
 QY 614 TFSAGSYRVGPQPGHKHSLSSGGSFPVAGVAGQDQYRPHSAQPADRGYDRPRAVSA 673
 Db 361 TFSAGSYRVGPQPGHKHSLSSGGSFPVAGVAGQDQYRPHSAQPADRGYDRPRAVSA 419
 QY 674 LATESGHPDSQKPTHTGTSDSYAPRDCITPLNQMTALL 715
 Db 420 L-DSRDPASQSQMT-SGDDGYSAPRNGLAPLNQMTALL 458
 RESULT 8
 US-09-759-130B-78
 ; Sequence 78, Application US/09759130B
 ; Publication No. US2003002279A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Millennium Pharmaceuticals, Inc.
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Fraser, Christopher C.
 ; APPLICANT: Sharp, John D.
 ; APPLICANT: Barnes, Thomas S.
 ; APPLICANT: Kirst, Susan J.
 ; APPLICANT: Mackay, Charles R.
 ; APPLICANT: Myers, Paul S.
 ; APPLICANT: Leiby, Kevin R.
 ; APPLICANT: Wrighton, Nicolas
 ; APPLICANT: Goodearl, Andrew
 ; APPLICANT: Holtzman, Douglas A.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
 ; FILE REFERENCE: M100-5350N1M
 ; CURRENT APPLICATION NUMBER: US/09/759,130B
 ; CURRENT FILING DATE: 2002-09-16
 ; PRIOR APPLICATION NUMBER: US 09/479,249
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: US 09/559,497
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: US 09/578,063

```
;; PRIOR FILING DATE: 2000-05-24
;; PRIOR APPLICATION NUMBER: US 09/333,159
;; PRIOR FILING DATE: 1999-06-14
;; PRIOR APPLICATION NUMBER: US 09/596,194
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 09/342,364
;; PRIOR FILING DATE: 1999-06-29
;; PRIOR APPLICATION NUMBER: US 09/608,452
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/393,996
;; PRIOR FILING DATE: 1999-09-10
;; PRIOR APPLICATION NUMBER: US 09/602,871
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: US 09/420,707
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 78
;; LENGTH: 235
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-759-1308-78
```

```
Query Match      33.68; Score 1265; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 2,7e-90;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 481 RKKKKGSPYSAAEQKTDCKMOKIYPPARHQAFTISYDNKEMTKLDLITSDMADY 540
Db 1 RKKKKGSPYSAAEQKTDCKMOKIYPPARHQAFTISYDNKEMTKLDLITSDMADY 60
OY 541 QOPLMIGTGYTRKSGTRFMDTDAEAGVSTDAAGHYDCQORAGRHEVALPLAPEPEY 600
Db 61 QOPLMIGTGYTRKSGTRFMDTDAEAGVSTDAAGHYDCQORAGRHEVALPLAPEPEY 120
OY 601 ATPYERHVALHAPHSAGSGYRVPGOPGKHKSLSGSGSPVAGVAGDGDYORPHSNOP 660
Db 121 ATPYERHVALHAPHSAGSGYRVPGOPGKHKSLSGSGSPVAGVAGDGDYORPHSNOP 180
OY 661 ADRGDRKAVASALATESGHPDSQKPTHPGTSDSYSAPRDLPLPLNOTAMTALL 715
Db 181 ADRGDRKAVASALATESGHPDSQKPTHPGTSDSYSAPRDLPLPLNOTAMTALL 235
```

```
RESULT 9
US-10-060-830-3
; Sequence 3, Application US/10060830
; Publication No. US20030032154A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
; FILE REFERENCE: PB0169
; CURRENT APPLICATION NUMBER: US/10/060,830
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/325,062
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 1123
; SOFTWARE: Neomica Sequence Listing Engine
```

```
;; SEQ ID NO 3
;; LENGTH: 729
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-060-830-3
```

```
Query Match      27.68; Score 1042; DB 9; Length 729;
Best Local Similarity 36.4%; Pred. No. 2,7e-72;
Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;
```

```
OY 20 LALLIYASAPRLIAAEELGDCGHLYTQDGGTMSKRYPCPTYPHNTCECTITVPKGR 79
Db 5 LLLLVLLLLLEDGAGQDGGHVTLPESGTLISNVPOTYPSVCEWEIRKMER 64
OY 80 LILRLGDDIE-SQTCADYLLFTS---SSDQYPCYG-SMTYKELLMTSEVYAFE 133
Db 65 VRIKRGDDIEDSDSCHNRYLRIYNGIVSNTEIGKYGLDLQNNHSLSEKGNETLLFM 124
OY 125 SGHVSGRGFLASYVIDKODLITCLDYSANFLPEFEFSKYPACCLLPFAISGTHPGY 184
OY 134 SGSHISGRGFLITYASSHPDLITCLERASHYLTKEYSKFCPCAGCRDVAAGDISGMVDGY 193
OY 194 RDTSLCKAAIHAGIINDELGGQISVLRKGISREGLANGVLSRDSLDKRLFTSN 253
Db 185 RDSPLCAVAGVAGVSNLTGQISVLSKGIPIYESSLANVYSVGHLSLFTERTS 244
OY 254 GCSRSLSFE---PDGQIRASSMSQSVNESGDQVHMSFGARLDQGPMSAGDSNNHK 309
Db 245 GCYGTGLNESGVIAIDPQITASSVLEMTDHTGQENSMKPKARKRKPGPMARATDE--- 301
OY 310 PREWLEIDLGEKKITGITRTGTSQSNFNYKSFVNFKNNSKWKTYKGIIVNEEKVF 369
Db 302 -YOWIQLDINKEKKTGITITGTSWEHNYVYSARILYSDGQKWTYRREPVEODKIF 360
OY 370 GNSNFRDPVONNFIPFIVARYRVYPOTWHORIALKYLIGCO-----ITQ----- 416
Db 361 QGNKDYHODVKNRNLPLIARIRVNPQOQOKIAMKHELGCOPITPKGRPKLQPPPP 420
OY 417 --GNDSLVWRKTSQTSYSTKKEDETTIRPI-----PSEET--STGINITT----- 458
Db 421 RNSND---LNTATAPRIAKGRAPKFTQPLQPSNSNEFPQOTEGTASPIDIRNTYVPN 476
OY 459 ----VAIPVLVLYVFGMGIFPAF-----RKKKKKS---PYCSAERQKTDCKQ 503
Db 477 VTKDVALAVALVPLVWLTLILLYLCAMHMRNKKKTEGTIDLPYV---DRAGMKG 532
OY 504 IKYFP-----ARHQAFTISYDNKEMTKLDLITSDMADYQOPLMIGTGYT 552
Db 533 MKQFLPAKAVDHEETPVKYSSE--VNHLSPREVT---TVLQADSALTAQPLVGIGYGL 587
OY 553 RKGSTFRPMDTDAEAG-VSTDAGGHYDCQORAGRHEVALPLAPEPEYATPIV---ERH 608
Db 588 HQBSTFKP--DEKRGVAGADDP---VNSPQEVYHAAVEPLITGPPEYATPIIDMSGH 642
OY 609 VLRHSTFSAQSGYRVPGOPGKHKSLSGSGSPVAGVAGDGDYORPHSNOPADRGYRP 668
Db 643 PTVSGQPSSTSTFKATGNP-----PPL--VGTWYTLISRTDSCSSAAQADYTP 689
OY 669 KA 670
Db 690 KA 691
```

```
RESULT 10
US-10-060-830-1114
; Sequence 1114, Application US/10060830
; Publication No. US20030032154A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
; FILE REFERENCE: PB0169
; CURRENT APPLICATION NUMBER: US/10/060,830
```

```

: CURRENT FILING DATE: 2002-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/325,062
: PRIOR FILING DATE: 2001-09-25
: NUMBER OF SEQ ID NOS: 1123
: SOFTWARE: Aeonica Sequence Listing Engine
: SEQ ID NO 1114
: LENGTH: 653
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-060-830-1114

```

```

Query Match      23.2%; Score 873.5; DB 9; Length 653;
Best Local Similarity 35.1%; Pred. No. 2,6e-59;
Matches 224; Conservative 102; Mismatches 211; Indels 101; Gaps 21;

```

```

QY 99 LFTSSDDQGYPCG-SMYPKELLNTSEYVAFESGSHSGFLLTASSHPDLIT 157
DB 13 LLEDAAOQKCYGGLQNHSSKSGNEITLLFMGSHVSGFLASYSVIDKODLIT 72
QY 158 CLERASHYLTXYKSKPCACRDVAGDISGMYGYRDTSLCKAIAHAGIADLGGOI 217
DB 73 CLDTASFTLEPERSKYRACCLLPFAISGTFIPGIVSDSPLCAGVHAGVSTLGGOI 132
QY 218 SVLQKGISRYEGLANGVLSRDGSLDKRFLFTSNCSRSLSFE---PDGQIRASSSW 273
DB 133 SVYISKGIPIYESSLANNVSVCHLSLSTLFTKSCYGLGMSGVINDPQTASSVL 192
QY 274 QSVNESGDQVHNSPGARLDODGPRSVASGSSNNHAKREWELEIDLGKKKITGRTTGST 333
DB 193 EMTDHTGQESMKPKKRLKPPGPMFAFATDE---YOMLQIDLNEKKITGITTTGST 248
QY 334 QSNFNFYKSFVAMFNKNNKMKTYKGIYVNEEKVFQGNFNDPVOONFPIPIVATYVR 393
DB 249 MVEHNIVYSARILYSDGQKATYVREPVEQDKIFQGNKDYHODVKNNEFLPIIATFIR 308
QY 394 VNPOTWRIALNVELIGCO-----ITQ-----GNDSLWRTSOSTSVSTKED 438
DB 309 VNPOTWRIALNVELIGCOFIPKGRPKLTOPPPRNSND---LKNTPAIPRIANGRA 364
QY 439 EITTRP-----PSEET---STGINTT-----VAIPVLVLLVLRAGMIFA 478
DB 365 PKFTQPLQPRSSNEFPQOTQTAIPDIRNTYTPVNTKDALAALVAVLVLWLTLLIL 424
QY 479 AF-----RKKKKKGS---PYGSAEAKTDCMKOIKYF-----ARHOSAEF 516
DB 425 ILVCAHMKRRKKKTETTYLPLTW---DRAGMKKMKOQLPAKAVDHETTPARYSSE- 479
QY 517 TISYDNKEKTKOILDTSMADYQOPLMTGTGTVTRKSGTFRPMOTDAEAG-VSTDAG 575
DB 480 -VNHLSTREYTV---TVQADSAEVAQPLVGIQVITLHQSTFKP--BEKREAGVADLP- 532
QY 576 GHVDCQPARHREYALPLAPREPEYATPIY---ERHVLRAHTPSAOGIYVPGQPHKH 632
DB 533 --TNSPQOEYVHAIAELPTTGPETATPILMDNSGHPPTSVGQPTSTFKAATGNP--- 586
QY 633 SLSSGGSFVAGVAGDGDYQRPHSADPADRGYDRPKA 670
DB 587 -----PPL--VGTYNTLSRDSKSSAAQOIDTPKA 615

```

```

RESULT 11
US-10-104-440-4
: Sequence 4; Application US/10104440
: Patent No. US20020132774A1
: GENERAL INFORMATION:
: APPLICANT: KLAGSBRUN, Michael
: APPLICANT: SOKER, Shay
: APPLICANT: MIAO, Hua Quan
: TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
: TITLE OF INVENTION: THEREOF
: FILE REFERENCE: 48802 C
: CURRENT APPLICATION NUMBER: US/10/104,440
: CURRENT FILING DATE: 2002-03-22
: PRIOR APPLICATION NUMBER: 09/580,803
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: 60/069,155
: PRIOR FILING DATE: 1997-12-09
: PRIOR APPLICATION NUMBER: 60/069,687
: PRIOR FILING DATE: 1997-12-29
: PRIOR APPLICATION NUMBER: 60/078,541
: PRIOR FILING DATE: 1998-03-19
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 931
: TYPE: PRT
: ORGANISM: human
US-10-104-440-4

```

```

Query Match      9.3%; Score 351.5; DB 12; Length 931;
Best Local Similarity 22.2%; Pred. No. 1.3e-18;
Matches 142; Conservative 99; Mismatches 255; Indels 143; Gaps 25;

```

```

QY 41 CGHLYTQDGSMTSKRYPTGTPPHNTCEKTIYPR-GKLLILRLG-DLIESQTCASDY 98
DB 28 CGGRLNSKDGAYTSPYPODPYSHONCEWIVYAFEPNOIVLNFNPFIEIKDKCYDR 87
QY 99 LTF-----TSSSDQGYPCGSMYVPEKELLMTSEYVAFESGSHSGFLLTY-----AS 149
DB 88 IEIHDGSEBNDLGRKCGNI-APPTIISGSMIYIFTSDYARQCGFSLRIEIFKTS 146
QY 150 SD-----HEDLYTC-----LERASHYLTXYKSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFEKYPHNDCFTTILAKPKMELIQLFLFDEHPLOYGEG 206
QY 174 -CPAGCQDVAGDIS--GNMVDGYNDSILCKAIAHAGIADLGGQISVLRQKGISRYEG 230
DB 207 DCKYDMDLMDGIPHYVGLIGKCYCGTTPSELSSSTGLSLTFTHTDAVAK-----DG 259
QY 231 ILANGVLSRDGSLDKRFLFTSNCSRSLSFE---PDGQIRASSMQSVNESGDQVHNS 286
DB 260 PSARYIYVHPEPLENFO-----CNVPLMGESGRINBDISASTY-----SDGRWT 305
QY 287 PGARLDODGPRSVASGSSNNHAKREWELEIDLGKKKITGIRTTG--STQSNFNYKSF 344
DB 306 PQOSRLHGDDNGWTPNLDNS---KEYLOYDLRLFTLFLAATOGAISRTONGYVKSX 361
QY 345 VAMFNKNNKMKTYKGIYVNEEKVFQGNFNDPVOONFPIPIVATYVRVYVPTWRIAL 404
DB 362 KLEVSTGEDMAYIRHKNH--KVFOANNDAATEVVLKRLAHLPTREVRIRPOTWHSIGA 419
QY 405 LKVELICQITQGNDSLVWTKTS---OSTSVSTRKKEDETTRP-----IP 446
DB 420 LRLELFGCRVTDAPCSNMLGMLSLINDSQISASSTOETVIMSSAARLVSSRSGMFPRIIP 479
QY 447 SEETSTGINTTVAIPVLVLLVLFVAGMG--IFA---AFRRKKKGSFYGSAEAKTDC 499
DB 480 QAGPGEEMLVQDICTPRTVKGVIIGARGGDSITAYVARAFVKKFKYSYSLNGKD----- 534
QY 500 CMKQIKTFFPARHSA-EFTISYD-----NEKEWTKOILDTSD 536
DB 535 -WEYIDPRTQOPLKFGNNHMDTPDIRRDPPIPAQVRYVPERWSPAGIGMLEVIGCD 593

```

QY 537 MADYQDPLMIGTGTTRKSGTFRPMDTDAEAGVSTDAG 575
 Db 594 WTD-SKPTVETLGTPTVSKSEETTPYPTTEE-----ATTECG 627

RESULT 12

US-10-104-440-2
 ; Sequence 2, Application US/10104440
 ; Patent No. US20020132774A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KLAGSBRUN, Michael
 ; APPLICANT: SOKER, Shay
 ; APPLICANT: MIAO, Hua Quan
 ; TITLE OF INVENTION: ANTAGONISTS OF NEUROFILIN RECEPTOR FUNCTION AND USE
 ; FILE REFERENCE: 48802 C
 ; CURRENT APPLICATION NUMBER: US/10/104,440
 ; CURRENT FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: 09/580, 803
 ; PRIOR FILING DATE: 2000-05-30
 ; PRIOR APPLICATION NUMBER: 60/069, 155
 ; PRIOR FILING DATE: 1997-12-09
 ; PRIOR APPLICATION NUMBER: 60/069, 687
 ; PRIOR FILING DATE: 1997-12-29
 ; PRIOR APPLICATION NUMBER: 60/078, 541
 ; PRIOR FILING DATE: 1998-03-19
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 923
 ; TYPE: PRT
 ; ORGANISM: human
 ; US-10-104-440-2

Query Match 8.9%; Score 335; DB 12; Length 923;
 Best Local Similarity 24.9%; Pred. No. 2.4e-17;
 Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

QY 17 RGL-----LALLNAVAPRLQAEELGDCGHLVYODSGTMTSKNYPGTYPNHTYCEK 70
 Db 3 RGLPLCALVALVLAAPGAR-----NDKCGPTIKIESGYLSPGYHSYPSKCEW 56
 QY 71 TTVVPGK-KRLILRLG-DLDIESQTCASDYL-LFTSSDO---YCPYCSMTVPKELLN 124
 Db 57 LIDAPDYQIRIMNFNFPLEDRDCKYDYVEVFDGENENGHRGKCGRI-APPVVS 115
 QY 125 TSEVTVFEESGSHSGRGLFYA-----SSDHDLLITC-- 158
 Db 116 GPFLFIFVSDYETHGAGFIRKIRFKRPECSONTTTSGVTKSPGFPEKYPNSLCTY 175
 QY 159 -----LERASHYLTKEYSKFCPCAG--CRDVAAGDISGMVVDGYRDTSLCKRAAH 205
 Db 176 IVFAPKMSIETLFEESDLEPDSNP--PGGMPCRYRDLER-----WQGFDP-----VGRP 223
 QY 206 AG-IIDLEGQI-----SVLQKRGISRYEISILANGVLSRDGSLSDKRFLE 250
 Db 224 IGRYCGKRTGRIRSSGILSMVFYTTDSAJAKGEGFSANYSV-SEDFK----- 274
 QY 251 TSNCCSASLSFEP---DQIRASSSMQSVNESGDVHWSFGARLQDQGPMSAGDSSN 306
 Db 275 -----CHEALMESGETHSDITASSQST-----NMSAERSRLNYPENGWTPBEDSY 322
 QY 307 NHRPREMLEDLGEKKKTIGRTTG--STQSNFNFYKSVNMFKNNSKWKYKGIYNN 364
 Db 323 -----REMIQVDCILIRVTAVGTGSAISKETKKYKTKKIDIVSSNGEWMITIK--EGN 376
 QY 365 EEKVFQGSNFRDPVONNFIPIVARYRVVPOTWQRIALAKVELIGCOI-----OGNDS 420
 Db 377 KPVLFQGNINPTVYVAVFEPKPLITREVRIKPAITWETGISMBREYVGCKITDPCSGMLG 436
 QY 421 LVWKTQSSTSVSTKKED 438

Db 437 MWSGLISDSQITSSNOGD 454

RESULT 13

US-10-115-563-14
 ; Sequence 14, Application US/10115563
 ; Publication No. US20030008307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffin, John H
 ; APPLICANT: Greengard, Judith S
 ; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
 ; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
 ; AND COMPOSITIONS THEREOF
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: The Scripps Research Institute, Office of
 ; Patent Counsel
 ; STREET: 10666 NO. US20030008307A1th Torrey Pines Road, TPC 8
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/115,563
 ; FILING DATE: 02-APR-2002
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/410,488
 ; FILING DATE: 24-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fitting, Thomas
 ; REGISTRATION NUMBER: 34,163
 ; REFERENCE/DOCKET NUMBER: 449.0
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-554-2937
 ; TELEFAX: 619-554-6312
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2224 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
 ; US-10-115-563-14

Query Match 8.4%; Score 317; DB 9; Length 2224;
 Best Local Similarity 42.0%; Pred. No. 2.2e-15;
 Matches 71; Conservative 30; Mismatches 52; Indels 16; Gaps 5;

QY 253 NCCSRLSEF-----PDGIRASS---SMQSVNESGDVHWSFGARLQDQGPMSAGDSS 305
 Db 2064 NCCSTPLGEMENKIRKQITASSFKSW----GD--YMEPFARLNAQGRVNAQAKA 2116
 QY 306 NHRPREMLEDLGEKKKTIGRTTG--STQSNFNFYKSVNMFKNNSKWKYKGIYNN 365
 Db 2117 NNK-QLLEIDLKIKITAITTOGCKSLSEMTYKSTTHYSQGVEMKPYRLKSMV 2174
 QY 366 EEKVFQGSNFRDPVONNFIPIVARYRVVPOTWQRIALAKVELIGCOI 414
 Db 2175 DKIFEGNTTKGHVKNFNPILISFRIVIPRTWQSTILRLLEGCID 2223

RESULT 14

US-10-095-718-4
 ; Sequence 4, Application US/10095718
 ; Patent No. US20020131956A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walsh, Christopher

APPLICANT: Chao, Hengjun
APPLICANT: Burslein, Halm
APPLICANT: Lynch, Carmel
APPLICANT: Stepan, Tony
APPLICANT: Munson, Keith
TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
FILE REFERENCE: 35052/204375
CURRENT APPLICATION NUMBER: US/10/095,718
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/689,430
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/158,780
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1431
TYPE: PRT
ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match 8.1%; Score 306.5; DB 12; Length 1431;
Best Local Similarity 40.5%; Pred. No. 7.5e-15;
Matches 68; Conservative 26; Mismatches 61; Indels 13; Gaps 4;

OY 253 NCSSRSLSFE---PDGQIRASSSMQSVNESGQVHSPGOARLDQGPMSAGSDSSNNH 308
DB 1271 NCSSMPLGMSKASDAQITASSYLSMLAT-----WSPQARLHIGRTNMRQANN- 1324
OY 309 KPWELEIDLGKKITGIRTTGTSQSNFNYKSVFNNFKNNNSKMTYKGIYNNEKY 368
DB 1325 -PREWLVQVFRKTKYTGITQGVKSLISMVYKEFLISSQDGHNTLF--LQNGKVKY 1381
OY 369 FQGNSEFRDVPQNNFIPPIVARYRVVPQTHQRIALKEVLCQITQ 416
DB 1382 FQGNRDSSTPVRNREPLVARYRLHPQSMHIALRLVLCQDTQ 1429

RESULT 15
US-10-190-593-2
Sequence 2, Application US/10190593
Publication No. US20030022221A1
GENERAL INFORMATION:
APPLICANT: LANGIT, Emanuel et al.
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
FILE REFERENCE: CL001246
CURRENT APPLICATION NUMBER: US/10/190,593
CURRENT FILING DATE: 2002-07-09
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 343
TYPE: PRT
ORGANISM: Human
US-10-190-593-2

Query Match 7.9%; Score 299; DB 9; Length 343;
Best Local Similarity 37.3%; Pred. No. 3.6e-15;
Matches 63; Conservative 33; Mismatches 55; Indels 18; Gaps 5;

OY 253 NCSSRSLSFE---PDGQIRASSSMQSVNESGQVH---WSPGOARLDQGPMSAGSD 303
DB 184 NCANPGLGKNNSTPDKQITASSSYKTWG-----LHLSMNPSTYARLDKQGNFNAWVAGS 238
OY 304 SSNNKRPREMLEIDLGKKITGIRTTGTSQSNFNYKSVFNNFKNNNSKMTYKGIYN 363
DB 239 YGND---QWLQVLDGSSKEVYTGILITQGARNGSVQFVASIKVAYSNDSANMTEYQDPRT 294
OY 364 NEKYVQGNSEFRDVPQNNFIPPIVARYRVVPQTHQRIALKEVLCQ 412

DB 295 GSSKIFPGMNDHSHKKNLEFPPIIARYVRLPVANHNRIALRLLELGC 343
Search completed: May 15, 2003, 13:21:32
Job time : 31.0681 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 20.546 seconds
(without alignments)
3345.478 Million cell updates/sec

Title: US-10-003-132-2

Sequence: 1 MYPGARGGAGALRAAGRGILL.....YSAPRDCLPLNQMTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pIR1:*
2: pIR2:*
3: pIR3:*
4: pIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	331	8.8	2211	1 KFB05	coagulation factor
2	330	8.8	927	1 J00948	A5 antigen precursor
3	321	8.5	2224	1 KFB05	coagulation factor
4	318	8.4	2183	2 T42764	coagulation factor
5	299	7.9	218	2 A47285	milk fat globule-p
6	297.5	7.9	427	2 JC4915	ags protein precursor
7	285.5	7.6	216	2 A44258	factor VIII-associated
8	285.5	7.6	2351	1 E2HU	coagulation factor
9	276.5	7.3	463	1 A36479	milk fat globule m
10	276.5	7.3	2319	2 A47004	coagulation factor
11	271.5	7.2	2133	2 T42763	coagulation factor
12	268	7.1	401	2 S65138	glycoprotein antigen
13	268	7.1	427	2 S74211	PAS-6/7 protein precursor
14	259.5	6.9	409	2 T11743	polyprotein - pig
15	217	5.8	845	2 JC5256	adipocyte transcri
16	216.5	5.7	719	2 S51739	transcription repr
17	192.5	5.1	1283	2 T11799	neurexin IV - fru1
18	190	5.0	3133	2 S52093	hemocytin - silkwo
19	186.5	4.9	686	1 A59271	Ra-reactive factor
20	174	4.6	1524	2 T30337	polyprotein - Afri
21	172.5	4.6	319	2 T51569	DVS-2 protein - Af
22	168	4.5	3623	2 T09456	intrinsic factor-B
23	165.5	4.4	986	1 B56788	procollagen C-endo
24	165	4.4	730	1 B56788	procollagen C-endo
25	165	4.4	823	1 A58788	procollagen C-endo
26	163.5	4.3	449	2 A55362	procollagen I C-pr
27	162.5	4.3	819	2 I48859	tyro 10 receptor k
28	160.5	4.3	694	2 JC5554	complement subcomp
29	158	4.2	991	2 I49540	procollagen C-endo

30	156	4.1	707	2 JC2218	procollagen C-endo
31	156	4.1	3623	2 T08618	intrinsic factor-B
32	155.5	4.1	855	2 T42621	protein-tyrosine k
33	153	4.1	402	2 JH0403	procollagen I C-pr
34	149.5	4.0	695	1 S05008	complement subcomp
35	142.5	3.8	1385	2 T14158	neurexin IV - mous
36	141	3.7	688	1 C1H05	complement subcomp
37	141	3.7	2083	2 T42721	CRP-ductin-alpha p
38	140.5	3.7	1081	2 T31083	paranodin - rat
39	137	3.6	705	1 C1H0RB	complement subcomp
40	136.5	3.6	699	1 I54763	Ra-reactive factor
41	136	3.6	1019	2 A38738	coagulation factor
42	135	3.6	737	2 T15615	hypothetical prote
43	133.5	3.5	1594	2 T30549	hensin - rabbit
44	133	3.5	855	2 JC7731	membrane-bound arg
45	133	3.5	855	2 JC7775	membrane type-seri

ALIGNMENTS

RESULT 1
KFB05
coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C:Accession: A42580; A36497
R:Guinot, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; M01D:92147638; PMID:1737753
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GUI>
A:Cross-References: GB:M81440; NID:9163037; PIDN:AAA30512.1; PID:9163038
A:Note: Sequence extracted from NCBI backbone (NCBIN:80774; NCBI:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:Title: Identification and characterization of a phospholipid-binding site of bovine
A:Reference number: A36497; M01D:91072354; PMID:2254316
A:Accession: A36497
A:Molecule type: Protein
A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584; 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'
Biochemistry 33, 13109-13116, 1994
A:Title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A55979; M01D:95034740; PMID:7947716
A:Contents: annotation
A:Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:Pathway: blood coagulation
A:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-28/Domain: signal sequence [status predicted <SIG>
F:29-221/Product: coagulation factor V #status predicted <MAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VHA>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-695/Domain: A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (Q-X-T/R-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F:1572-1992/Domain: A3 <DA3>
F:1572-1992/Region: ferroxidase repeat homology <FO3>
F:1654-1152/Region: phospholipid binding #status predicted
F:1893-2051/Domain: C1 <DC1>
F:1893-2048/Domain: discoidin I amino-terminal homology <DNI>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discoidin I amino-terminal homology <DNI>
F:167-193, 248-329, 499-525/Disulfide bonds: #status experimental

A:Note: parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortel, T.L.; Quinn-AlLEN, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; M0ID:95210278; PMID:7696276
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:119896; OMIM:227400
A:Map position: 1923-1923
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 653/2
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa protease.
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal; ferroxidase reductase; protein C; protein S
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; protein C; protein S
F:1-48/Domain: signal sequence #status predicted <SIG>
F:219-2224/Product: coagulation factor V #status predicted <MAT>
F:229-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:229-345/Domain: A1 <DA1>
F:333-329/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:1906-2061/Domain: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F:51.55.239.297.460.468.554.741.752.760.776.782.821.938.977.1074.1083.1103.1106.1479.1493.1574.1573.248.329.500-526.603-684.1725-1751.1907-2061.2066-2221/Disulfide bonds: #status predicted
F:334-333/Cleavage site: Arg-San (protein C) #status predicted
F:363,693,1346/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382,1338/Binding site: carbohydrate (Asn) (covalent) #status absent
F:534-533/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match. 8.5%; Score 321; DB 1; Length 2224;
Best Local Similarity 42.6%; Pred. No. 3.9e-14;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

OY 253 NCGSRSLSE-----PQOIRASS---SMQSVNSESDDYHMSGCARLDQGGSMAGDSS 305
DB 2064 NCGSTPLGHNKGIENKQIRASSFKKSW-----GD--YWEFFARLNLNQGVRVNAQARA 2116
OY 306 NNHKPREMLEIDGKKKRTGIRGTGSDQSNFNFVSKFVNFKNNKSKKRYKGIYVNE 365
DB 2117 NNKK--QWLEIDLLKTKIKTAVITLQGGCKSLSEMYKSYTTHISGGVEMKRYRLKSSWV 2174
OY 366 EKVFGQNSNFRDYPVNNFPIPIYARVVRVQTHQRIALAKVELIGCOT 414
DB 2175 DRIFEGNTNKGHVKNFNPILSRIFRIVPIKTNQSTALRIELFGCDI 2223

RESULT 4
T42764
coagulation factor V - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Mousasalil, M.; Kaufman, R.J.; Ginsburg, B.
Biology 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by protein C
A:Reference number: Z22270; M0ID:98282202; PMID:9616155
A:Accession: T42764

A:Status: prelminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-2183 <TAN>
A:Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC99553.1
C:Function:
C:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasmin
E:350-662/Domaln: ferroxidase repeat homology <FOX1>
E:1541-1864/Domaln: ferroxidase repeat homology <FOX2>

Query Match	8.4%	Score 318;	DB 2;	Length 2183;
Best Local Similarity	40.8%	Pred. No. 6.2e-14;		
Matches 69;	Conservative 31;	Mismatches 53;	Indels 16;	Gaps 5

QY	253	NGCGRSLSTFE-----PDCQIRANS-----MOSVNSGDOVHNSPCQARLODQPEPMASGDSS	305
Db	2023	NGCSTPLGLEDRGRQDQIRANSRKSWM-----GD--VYEPPLATLNQGRNMAQAQA	2077
QY	306	NNHPRMELDELGEKKKTIIGRTGTSQSNENFYKSPFMANRKNNSKKYTKIGLVNNE	365
Db	2076	NNNN--OMLQDLTKIKVATLQYGGKSLSEMYKYSVSIQYSDGVAMKPPRQKSSNV	213

QY 366 EKVEGNSNEFRDPVONNETPPIVARYVRVPQTHORIALKVELLIGCQI 414
:::||||: ::| |||:::||||: |||::| |||
Db 2134 DKIEGNSNTKGHMKNFNPPIISRIRIIPKTFWQSIARLELFGCDI 2182

RESULT 5
AA7285
milk fat globule protein - human (fragment)

C:Accession: A47285
R:larocca, D.; Peterson, J.A.; Urra, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991

A:Reference number: A47285; MWID:91371351; PMID:1909332
A:Accession: A47285
A>Status: preliminary

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F1:56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F2:60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match	7.9%;	Score 299;	DB: 2;	Length 218;
Best Local Similarity	37.38%;	Pred. No. 4.5e-14;		
Matches 63; Conservative	33;	Mismatches 55;	Indels 18;	Gaps 5

```
QY      253 NGCGRSLFE---PDGQIRASSSSWOSYNESGDVH--WSPGARLQDDG--PSWASGD 303
        ||| : | : |||| :: : | ||| || : : |
DB      59 NCGANPLGLKNNSIDPKQITASSSYKTWG-----LHLFSNPPSYARLDKQGNFNNAVAGS 113
```

```

0Y      304 SSNNHKKPREWLEIDGKKKKTGIRTGTOSTQSFNFYKSFVMMKNKNSKTKTYGIYN 363
      1:  ::::| |::| |  :  ::| ::| ::| |  :
Db      114 YGND----QMLQVDGSSKKEVGTITGARNFSSVQFVASYKVAVSNDASAMWTEYQDPR 169

```

```

Oy      364 NEEKVFOGNSNFRDPVONNEIPPIVARYRVVPQTWHORIALKVELIGC 412
          |::||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      170 GSKIFPGNWDNHSKKNLFETPIIARYRILPVAMHNRIALRELLGC 218

```

RESULT 6

ags protein precursor - rat
N:Alternate names: O-acetyl-Gd3 ganglioside
C:Synonyms: Batus normalis (Morris rat)

C;species: nactus norvegicus (norway rat)
C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000

R: Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996

[illegible]

Oy 369 KPELELDLGEKKRTITGTFSTQSPNEFFKSEFVNFKNNSKMATYGIYINNEKY 368
|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 110 -PKEMLOVDFOKTAKVTGTVOGVKSLLTSMATVEEFLISSODGHQTLP--PONGKAYV 166

Oy 369 FOGNSENFRDPVONNEFIPIVARVRVPDPTWRIALKEVLIGCQ 413
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 167 FOGNDSETFPVNSLDDPELLTRYLTHIPDSWWHQTALREVIAGE 211

RESULT 8
EZHU
coagulation factor VIII precursor [validated] - human
M.ALternate names: antihemophilic factor A; coagulation factor VIIC; procoagulant co-
J.Species: Homo sapiens (man)
C.Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C.Accession: J34318; A00525; I58059; A23584; A20174; A42348; A43986; S63527; S66445;
R.Glttschler, J.; Wood, W.I.
Hum. Mol. Genet., 1, 199-200, 1992
A>Title: Sequence of the exon-containing regions of the human factor VIII gene.
A.Reference number: I54318; MUID:93265012; PMID:1303178
A.Accession: I54318
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A.Residues: 1-1921,'S',1923-2351 <RES>
A:Cross-references: GB:M8648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R.Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Glttschler, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A>Title: Expression of active human factor VIII from recombinant DNA clones.
A.Reference number: A00525; MUID:85061548; PMID:6438526
A.Accession: A00525
A:Molecule type: mRNA
A.Residues: 1-2351 <MOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R.Toole, J.U.; Knopf, J.L.; Wozney, J.M.; Soltzman, L.A.; Buecker, J.L.; Pittman, D.D.
s. D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A>Title: Molecular cloning of a cDNA encoding human antithrombophilic*factor.
A.Reference number: I58059; MUID:85061550; PMID:6438528
A.Accession: I58059
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A.Residues: 1-74,'V',76-1259,'E',1261-2351 <RES>
A:Cross-references: GB:X01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R.Tiretti, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dhna, D.; Hartog, K.; K
.B.; Randolph, A.; Under, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J..
DNA 4, 333-349, 1985
A>Title: Characterization of the polypeptide composition of human factor VIII:C and t
A.Reference number: A23584; MUID:86081164; PMID:3935400
A.Accession: A23584
A:Molecule type: mRNA
A.Residues: 1-2351 <TRU>
A:Cross-references: GB:MA113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R.Eaton, D.; Rodriguez, H.; Venar, G.A.
Biochemistry 23, 505-512, 1986
A>Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A.Reference number: A26174; MUID:86159740; PMID:3082357
A.Accession: A26174
A:Molecule type: protein
A.Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <PA
R.Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A>Title: Identification and functional importance of tyrosine sulfate residues within
A.Reference number: A42348; MUID:92207952; PMID:1554716
A.Accession: A42348
A:Molecule type: protein
A.Residues: 20-36;355-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709
A.Experimental source: recombinant material from Chinese hamster ovary cells
A>Note: sequence extracted from NCBI backbone and corrected to correspond with the pu
r.Fay, P.U.; Smudzin, T.M.
. Biol. Chem. 264, 14005-14010, 1989
A>Title: Intersubunit fluorescence energy transfer in human factor VIII.

A:Reference number: A43986; MUID:89340500; PMID:2503509
 A:Accession: A43986
 A:Molecule type: protein
 A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <FAY>
 R:leyve, A.; van Schijndel, H.B.; Niehrs, C.; Huttnner, W.B.; Verbeet, M.P.; Mertens, K.; J. Biol. Chem. 266, 740-746, 1991
 A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for A:Contents: annotation; sulfation
 A:Reference number: A56109; MUID:91093266; PMID:1898735
 R:Gitchler, J.; Wood, W.I.; Goralika, T.M.; Wlton, K.L.; Chen, E.Y.; Eaton, D.H.; Venhar, Nature 332, 326-330, 1984
 A:Title: Characterization of the human factor VIII gene.
 A:Reference number: A56196; MUID:85061547; PMID:6438525
 A:Contents: annotation; Introns
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M. Protein Sci. 4, 740-746, 1995
 A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of A:Reference number: A56216; MUID:95338127; PMID:7613471
 A:Contents: annotation; disulfide bonds
 A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
 R:Fjalke, M.; Hedling, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M. Eur. J. Biochem. 234, 773-779, 1995
 A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; MUID:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752;753-759 <KJA>
 R:ind, P.; Larsson, K.; Spire, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg Eur. J. Biochem. 237, 19-27, 1995
 A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction A:Reference number: S66445; MUID:96048024; PMID:7556150
 A:Accession: S66445
 A:Molecule type: preliminary
 A:Status: preliminary
 A:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure to C:Genetics: GDB:119124; OMIM:306700
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro A:Pathway: Blood coagulation
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla F:1-19/Domains: signal sequence #status predicted <SIG>
 F:20-335/Product: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>
 F:20-356/Domains: A1 <DA1>
 F:23-348/Domains: ferroxidase repeat homology <FO1>
 F:392-759/Domains: A2 <DA2>
 F:402-730/Domains: ferroxidase repeat homology <FO2>
 F:1668-2351/Domains: B <DB0>
 F:760-1667/Domains: coagulation factor VIIIA light chain #status experimental <AC1>
 F:1709-2038/Domains: A3 <DA3>
 F:1716-2038/Domains: ferroxidase repeat homology <FO3>
 F:2039-2191/Domains: C1 <DC1>
 F:2039-2188/Domains: discoidin I amino-terminal homology <DN1>
 F:2192-2351/Domains: C2 <DC2>
 F:2192-2345/Domains: discoidin I amino-terminal homology <DN2>
 F:60, 258, 601, 776, 803, 847, 919, 962, 982, 1020, 1074, 1085, 1204, 1274, 1278, 1301, 1319, 1403, F:167-198, 267-348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/Disulfide bonds: #status F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365-737/738, 742, 1683, 1699/Binding site: sulfate (Tyr) (covalent) #status experimental
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:414, 442/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1667-1668/Cleavage site: Arg-Glu (undefined proteolysis) #status experimental
 F:1709-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2193-2345/Disulfide bonds: #status predicted

Query Match 7.6%; Score 285.5; DB 1; Length 2351;
 Best Local Similarity 37.6%; Pred. No. 1,5e-11;
 Matches 62; Conservative 32; Mismatches 58; Indels 13; Gaps 4;
 253 NGSRSLSFE-----PQOIRASSSMOSVNSGQVHSPQOARLQOQSPMASGDSNNH 308
 2191 NCSMPGMSKSAISDQITASSFTWFAV-----WSPKARLHLOGRNNARPQVNN- 2244
 309 KPREMLEIDGKKRTIGITGTGTSQNFNFYKSPFVNFKNNSKWTYKGIYNNDEKY 368
 2245 -PREMLQVDQKTKMYKTYGTGKSLTSMYKEFLISSQGHQWTFE--FQNGKVKY 2301
 369 FQGNSEFDPVQNNFPIVARYRVVQVTHORIALKVELICQ 413
 2302 FQGNDSFTPVNSLDPILLTRILRIHQSGVHQAIALRMEVLCE 2346
 RESULT 9
 A36479
 milk fat globule membrane protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A36479
 R:Scudis, J.D.; Dekutis, C.; Singer, K.L.; But, A.; Yuzuki, D.; Strinivasan, U.; Parry Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
 A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the A:Reference number: A36479; MUID:91046008; PMID:2122462
 A:Accession: A36479
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-463 <RTD>
 A:Cross-references: GB:M38337; NID:9199142; PIDN:AAA9534.1; PID:9199143
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom F:28-60/Domains: EGF homology <EG1>
 F:68-107/Domains: EGF homology <EG2>
 F:147-303/Domains: discoidin I amino-terminal homology <DN1>
 F:307-463/Domains: discoidin I amino-terminal homology <DN2>
 Query Match 7.3%; Score 276.5; DB 1; Length 463;
 Best Local Similarity 23.4%; Pred. No. 5.8e-12;
 Matches 112; Conservative 59; Mismatches 118; Indels 189; Gaps 21;
 94 CASDYLLFTSSSDQYGYCGSMTPVPELLNTSEVTVRFSGSHISGR-----GP 143
 16 CASG--LEAASGD-----FCDS-----SLCLN-----GGTCLGQDNDIYCLDEGF 55
 144 LLYTAS-----SDHDLITCLERASHLYLTKTEYKRCPCACRQVADISGM- 189
 56 TGLVCNTERGSPSPNCPVNDACLVY-LDTORGDIPTFYICQPVGYSGICHCETETNY 114
 190 -VDG-YRDTSLCKAAI-----HAGITVD----- 211
 115 NLGGEWFTTAVPTAVPAPRPDLSNNLASKCSQQLGHEGAILDSQISASTYTMGEW 174
 212 ---ELGQSLVLRKGI-----SRVEGIIANGVLSRDG----- 241
 175 GLGRMGPFLARLRYTGIVNMMHASNDSKRPVQVNLKRKRVSGVMTQGA-SRAGAEVY 233
 242 -----SLSDKRFPLFTSN-----GCSR- 257
 234 KTFKVVAYSLDGRKFEFTQDSGDKFEFLGNLNNNSIKVNNFNTLEAQYIRLYPVSCHRG 293
 258 -SLSEF-----PDGQIRASSSMOSVNSGDDVHNSPGGARLQDQ 296
 294 CTRFELGCELHGCLEPLGLKNNTPDSSQSSSTKTNLRA--FGWTPHLGRLDNQG 351
 297 P--SWASGDSNNHKKPREWLEIDLGERKKKTGRTGTSQNFNFYKSPFVNFKNNSK 354
 352 KINAW-----TAQNSNAKEMLVQDVGQROVTTGITGARGDFHQIYVESKVAHSDQVG 407
 355 WKTYKGIYNNDEKYFQGNSEFDPVQNNFPIVARYRVVQVTHORIALKVELICG 412

Db 408 WTVTE--E0GSSKVFQGLNDNNHKKNIKPEKPNARVRLVPSWNRITLRLLELGC 463

RESULT 10

A47004

C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999

C:Accession: A47004

R:Eider, B.; Laskich, D.; Gletscher, J.

Genomics 16, 374-379, 1993

A:Title: Sequence of the murine factor VIII CDNA.

A:Reference number: A47004; MUID:93300511; PMID:8314577

A:Accession: A47004

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2319 <EID>

A:Cross-references: GB:105573; NID:g192456; PIDN:AAA37385.1; PID:g192457

C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase

F:1-19/Domain: signal sequence #status predicted <SIG>

F:23-349/Domain: ferroxidase repeat homology <FO1>

F:402-730/Domain: ferroxidase repeat homology <FO2>

F:1686-2006/Domain: ferroxidase repeat homology <FO3>

F:2007-2156/Domain: discolidin I amino-terminal homology <DN1>

F:2160-2319/Domain: discolidin I amino-terminal homology <DN2>

Query Match

Best Local Similarity 7.3%; Score 276.5; DB 2; Length 2319;
Matches 61; Conservative 33; Mismatches 61; Indels 13; Gaps 4;

Y 253 NGCSRSLSFE---PDGQIRASSSMQSYNEGDDVHNSPGARLQDGPMSAGDSNNH 308

Db 2159 NSCSIPLEGMEKRVISDQITRTSSFTNFAF---WSPSOARLHLDRTNAMPQVND- 2212

Y 309 KPREMLEIDLGEEKKITGIRGTGSTQSNFNFYKSFVNFKNNNSKMTYKGIYVNEEKV 368

Db 2213 -PKWMLQVLDQKTKMKVGIITQGVKSLFTSMFVKEFLISSQDGHMWTQI--LVNGKRV 2269

Y 369 FQGSNSFRDPVQNNFIPPIVARYRVVPTQWRIALKEVLCGQITQ 416

Db 2270 FQGNDSSTPMNSLIDPLRLTRLRHPIHQIWEHIALRLLEITLGEAQ 2317

RESULT 11

T42763

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

C:Accession: T42763

R:Idlilar, P.

Submitted to the EMBL Data Library, August 1996

A:Reference number: 22269

A:Accession: T42763

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2133 <LOI>

A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AA06705.1

C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase

C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-2133/Domain: coagulation factor VIII #status predicted <MAT>

F:23-349/Domain: ferroxidase repeat homology <FOX1>

F:402-730/Domain: ferroxidase repeat homology <FOX2>

F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match

Best Local Similarity 7.2%; Score 271.5; DB 2; Length 2133;
Matches 61; Conservative 29; Mismatches 57; Indels 23; Gaps 5;

Y 253 NGCSRSLSFE---PDGQIRASSSMQSYNEGDDVHNSPGARLQDGPMSAGDSNNH 308

Db 1973 NSCSIPLEGMEKRVISDQITRTSSFTNFAF---WSPSOARLHLDRTNAMPQVND- 2212

Y 309 KPREMLEIDLGEEKKITGIRGTGSTQSNFNFYKSFVNFKNNNSKMTYKGIYVNEEKV 368

Db 2213 -PKWMLQVLDQKTKMKVGIITQGVKSLFTSMFVKEFLISSQDGHMWTQI--LVNGKRV 2269

Y 369 FQGSNSFRDPVQNNFIPPIVARYRVVPTQWRIALKEVLCGQITQ 416

Db 2270 FQGNDSSTPMNSLIDPLRLTRLRHPIHQIWEHIALRLLEITLGEAQ 2317

Y 309 KPR---EWLEIDLGEEKKITGIRGTGSTQSNFNFYKSFVNFKNNNSKMTYKGIYV 363

Db 2021 RPRVSSAEMLQVLDQKTKMKVGIITQGVKSLFTSMFVKEFLISSQDGHMWTQI--LV 2078

Y 364 NEEKYFQGSNSFRDPVQNNFIPPIVARYRVVPTQWRIALKEVLCGQ 413

Db 2079 GHTKVFQGNDSSTPMNSLIDPLRLTRLRHPIHQIWEHIALRLLEITLGEAQ 2128

RESULT 12

655138

C:Species: Bos primigenius taurus (cattle)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998

C:Accession: S65138

R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <AOX>

R:Maier, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-

I-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A:Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <MAT>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discolidin I amino-terminal homology; EGF hom

C:Keywords: glycoprotein

F:1-32/Domain: EGF homology <EG1>

F:40-79/Domain: EGF homology <EG2>

F:82-239/Domain: discolidin I amino-terminal homology <DN1>

F:243-401/Domain: discolidin I amino-terminal homology <DN2>

Query Match

Best Local Similarity 7.1%; Score 268; DB 2; Length 401;
Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

Y 253 NGCSRSLSFE---PDGQIRASSSMQSYNEGDDVHNSPGARLQDGP--SWASGSSN 306

Db 242 NGCTEPGLKNTIPKQITRTSSFTNFAF---WSPSOARLHLDRTNAMPQVND- 2212

Y 307 NHEKPREMLEIDLGEEKKITGIRGTGSTQSNFNFYKSFVNFKNNNSKMTYKGIYVNEEKV 366

Db 300 S---EWLEIDLGEEKKITGIRGTGSTQSNFNFYKSFVNFKNNNSKMTYKGIYVNEEKV 366

Y 367 KVFQGSNSFRDPVQNNFIPPIVARYRVVPTQWRIALKEVLCG 412

Db 356 KIFQGNDSSTPMNSLIDPLRLTRLRHPIHQIWEHIALRLLEITLGEAQ 401

RESULT 13

S74211

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000

C:Accession: S74211; S76114; S24181

R:Harrigard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat g

A:Reference number: S74211; MUID:97008954; PMID:8856064

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <HYA>

A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779

A:Accession: S78114
 A:Molecule type: protein
 A:Residues: 19-55;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
 R:Kim, D.H.; Kanno, C.; Mizokami, Y.
 Biochim. Biophys. Acta 1132, 203-211, 1992
 A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
 A:Reference number: S23926; M0ID:92353107; PMID:1643094
 A:Accession: S24181
 A:Molecule type: protein
 A:Residues: 383-394 <KIM>
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
 C:Keywords: blocked amino- end; disulfide bond; glycoprotein; milk
 F:1-16/Domain: signal sequence #status predicted <Sig>
 F:19-427/Product: PAS-6/7 protein #status experimental <Mat>
 F:24-58/Domain: EGF homology <EG1>
 F:66-105/Domain: EGF homology <EG2>
 F:108-265/Domain: discoidin I amino-terminal homology <DN1>
 F:269-427/Domain: discoidin I amino-terminal homology <DN2>
 F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
 F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
 F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:109-265;252-286;270-427/Disulfide bonds: #status experimental

Query Match 7.18; Score 268; DB 2; Length 427;
 Best Local Similarity 36.7%; Pred. No. 2.1e-11;
 Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

OY 253 NGCSRLSEF---PDGIRASSQSVNESGQVHMSPGQALDQGP--SWASGDSN 306
 DB 268 NGCTPEPLGKNDTIPNKQITASSYTWGLSA--FSWFPYARLDNGCKENATQAQNSA 325
 OY 307 NHRKPREMLDLEKKRTGRTGTSQSNFNFYKSFVNFKNNSKMTYGIYVNEE 366
 DB 326 S---EWLQDLGSGQRVYGIITQGARDFGHIQYVAARVAGDDGVTWETEDPGASES 381
 OY 367 KYFGNSNFRDPVQNNFPIPIVARYRVVPTQTHORIALKVELIGC 412
 DB 382 KIFPGNNDNNSHKKNIFEPFQARFRIQEVAVMHNRIITLVELLGC 427

RESULT 14

T11743
 P47 protein - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T11743
 R:Enslin, M.; Vogel, T.; Calvete, J.J.; Whole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
 Biol. Reprod. 58, 1057-1064, 1998
 A:Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zc
 A:Reference number: Z17325; M0ID:98206817; PMID:9546740
 A:Accession: T11743
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-409 <ENS>
 A:Cross-References: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
 A:Experimental source: testis
 C:Function:
 A:Description: may be involved in membrane remodeling and/or function as a zona pellucida
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:6-40/Domain: EGF homology <EGF>

Query Match 6.9%; Score 259.5; DB 2; Length 409;
 Best Local Similarity 25.0%; Pred. No. 7.9e-11;
 Matches 87; Conservative 46; Mismatches 102; Indels 113; Gaps 11;

OY 168 TYSKRCPCAGCRVADISGNMDGTRDTSL-----LCKAIIHGIIT-----ADELGC--- 215
 DB 72 TETICPCPGYTGIIHCEIICNAPLGKMGTAIDFOISASSMHLGFMGLQRMABELARLHR 131
 OY 216 -----GISVLQKRGISRYEGIIANGVLSRDGSL----- 243
 DB 132 AGIVNMTASNTDRNPIQVNLRLR---KAVTVVIOGA--SRAGSAEYKTFKFAVNSTDG 187

OY 244 -----SDKREL-----FTSMGC 255
 DB 188 RKQFIQGAESDQKIFMGLNDLSGLKVNLFEPLEQVYRVLPICHRCTARFELGCG 247
 OY 256 SRSLSFEPDG-----QIRASSSQSVNESGQVHMSPGQALDQGP--SWASGDS 304
 DB 248 ELSCAEPILGKNDTIPNKQITASSYTWGLSA--FSWFPYARLDNGCKENATQAQNSA 305
 OY 305 NHRKPREMLDLEKKRTGRTGTSQSNFNFYKSFVNFKNNSKMTYGIYVNEE 364
 DB 306 SAS---EWLQDLGSGQRVYGIITQGARDFGHIQYVAARVAGDDGVTWETEDPGASES 381
 OY 365 EEFVQNSNFRDPVQNNFPIPIVARYRVVPTQTHORIALKVELIGC 412
 DB 362 EGIIFGNDNNSHKKNIFEPFQARFRIQEVAVMHNRIITLVELLGC 409

RESULT 15

JCS256
 adipocyte transcription factor, AEBP1 - human
 C:Species: Homo sapiens (man)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: JCS256
 R:Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,
 Biochem. Biophys. Res. Commun. 228, 411-414, 1996
 A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expr
 A:Reference number: JCS256; M0ID:97079196; PMID:8920928
 A:Accession: JCS256
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-845 <OHN>
 A:Cross-References: DBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Query Match 5.8%; Score 217; DB 2; Length 845;

Best Local Similarity 25.7%; Pred. No. 2.6e-07;
 Matches 55; Conservative 40; Mismatches 97; Indels 22; Gaps 5;

OY 264 DGOIRASSSQ--SVNESGQVHMSPGQALDQGPVWASGDSNNHNPRLDLEKK 322
 DB 84 DNGIRASSMRLHGLGARGLNQGTATEDDYDGMACADDAR-----QWIEVDTRRT 139
 OY 323 KITGIRTGSTQSNFNFYKSFVNFKNNSKMTYGIYVNEE 366
 DB 140 RFGVLTQGRDSSIHDFVTFFVFGSNDQVWMTN--GYEEMTFHGNVDKTPVLS 197
 OY 383 FIPPIVARYRVVPTQTHORIALKVELIGCQITQ-----GNDSLVNR-----KTS 427
 DB 198 LPEPVARFRIPLTWNGSLCRLEVLGCVAPVSYAQNNEVATDDLDPRHSHYKDM 257
 OY 428 QSTSVSTKREDETTRIPSEETSGINTVVAI 461
 DB 258 RQLMKVNECPITTRISLGRSSRLKITAMEI 291

Search completed: May 15, 2003, 13:16:51
 Job time : 36.546 secs

THIS PAGE BLANK (USPTO)

STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-24

Query Match 9.4%; Score 354.5; DB 3; Length 906;
Best Local Similarity 21.5%; Pred. No. 2,4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

QY 41 CGHLYTQDSGTMSTKNTPTGTYPHNYCEKTIYTPK-GKRLILRLG-DLIDESQTCASDY 98
DB 28 CGGRPNKSDAGYITSPYDYPHONCEWIVAPENOKIVLNFNPFIEKHDKYDF 87
QY 99 LLF-----TSSSDQGYPCGSMWPKVREKLLNTSEVTVPEESGSHSGRGLTTY-----AS 149
DB 88 IEIRDGSEADLGRKCGNI-APPTIISGSVYIYFTSDYANOGAGFSILRYEIFRTGS 146
QY 150 SD-----HPLITC-----LERASHYLKTEYSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFEKYPHNDCFTTILAKPRMEIILQFLFDEHDPLOYGEG 206
QY 174 -CPAGCDVAGDIS--GNMDGYDTSILCKAAIHAGIADLGGQISVLOKRSIKRY-- 228
DB 207 DCKYDMDIDWGIPIHVGPLIGKTCGTPSKRLRSSTGILSTFTTDAVAVAKDGFSAARYL 266
QY 229 -----EGLANGVLSRDGSLSDKRFLLFTSNCGSRLSFEPPGQIRAS 270
DB 267 IHQEPENFOCNVPLGESHRIANEQISASTFSDGR----- 303
QY 271 SSMQSVNESGDQVHMSFGQARLADQGFPMASGSSNNHKKPREMLEIDLGEKKKITGIRTT 330
DB 304 -----WTPQGRLEHGDNDGWTPTNLDN-----KEYLOVDFRFLMTALATQ 345
QY 331 G---STOSNENFYVKSFMNKRNNNSKKTKYKGIYNNNEKYPOGNSNRDPQONFFPIPIV 388
DB 346 GAISREFOKGIYVKSYLEVSTNGEDMWVYRHGKNH-KIPIQANDTAEVVLNHLHPL 403
QY 389 ARYRVVTPQTHORIALKVELIGCOITQGNDSLWVRKTS---OSTSVSTKKEDEITRPI 445
DB 404 TRFIRIRPQTHWLGIALRLLEFGCRVTDAPCSNNLGLSLGIADTQISASTREIYLSPS 463
QY 446 -----PSEETSTGINTTVAIPIVLVLVLPAGMGIFA---AFRK 482
DB 464 AARLVSRSGWFPNPOAGGEEW-LQVLDLCTPKTVKGVIIQAGRGDSITAVBARAFVR 522
QY 483 KKKKGSYGSLEAKQTCCKQIKYFPAHQSA-EFTISYD----- 521
DB 523 KFKVSYSLNGD-----WEYIDDPRTQOYTKLFGNNHMYDTPDIIRFDPVPAQYRVRYPE 576

QY 522 --NEKEWTKDLITSDMADYQOPL-MIGTGYTRKSGSTFPMQTDABEAG 569
DB 577 RMSPAGIGMRLEVIGCMWTDKSPVETLIGPVYKSEFTTTPRMDADATFEG 627

RESULT 4
US-08-936-135-8
Sequence 8, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-8

Query Match 9.4%; Score 354.5; DB 3; Length 909;
Best Local Similarity 21.5%; Pred. No. 2,4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

QY 41 CGHLYTQDSGTMSTKNTPTGTYPHNYCEKTIYTPK-GKRLILRLG-DLIDESQTCASDY 98
DB 28 CGGRPNKSDAGYITSPYDYPHONCEWIVAPENOKIVLNFNPFIEKHDKYDF 87
QY 99 LLF-----TSSSDQGYPCGSMWPKVREKLLNTSEVTVPEESGSHSGRGLTTY-----AS 149
DB 88 IEIRDGSEADLGRKCGNI-APPTIISGSVYIYFTSDYANOGAGFSILRYEIFRTGS 146
QY 150 SD-----HPLITC-----LERASHYLKTEYSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFEKYPHNDCFTTILAKPRMEIILQFLFDEHDPLOYGEG 206
QY 174 -CPAGCDVAGDIS--GNMDGYDTSILCKAAIHAGIADLGGQISVLOKRSIKRY-- 228
DB 207 DCKYDMDIDWGIPIHVGPLIGKTCGTPSKRLRSSTGILSTFTTDAVAVAKDGFSAARYL 266
QY 229 -----EGLANGVLSRDGSLSDKRFLLFTSNCGSRLSFEPPGQIRAS 270
DB 267 IHQEPENFOCNVPLGESHRIANEQISASTFSDGR----- 303
QY 271 SSMQSVNESGDQVHMSFGQARLADQGFPMASGSSNNHKKPREMLEIDLGEKKKITGIRTT 330

```

Db 304 -----WTPOOSRLHGDNGMTPNLDN---KEYLOVDLRLFLMLTAIATQ 345
Oy 331 G--STOSNENFYKGFVAMNFKNNNSKMTYKGIYVNEKEVFGNSFNDPVOONNFIPIY 388
Db 346 GAISETOKGYIVKSKYLEVSTNGEDMAYYRHKNH--KIFQANDATEVYLNKLMPL 403
Oy 389 ARYVAVPQTMHQRALAEVLIQCOITQGNDSLVRKTS---QSTSVSTKKEDETIPI 445
Db 404 TRFIRPOTMILGIALRLLEFGCHVTAPCSNMLGMLSGLIADQISASTREYIMSPS 463
Oy 446 -----PSEETSGINITVAIPVLVLYVYAGMIFA---APRK 482
Db 464 AARLVSSSGWFPRNPQAPGEEW-LQYDLGTPKTVKGVIIQANGGDSITAVEARAFVR 522
Oy 483 KKKKSPYGSAAEAKTDCKQIKYEPFARHQA-EFTISYD-----521
Db 523 KFKVSYSLNGKD-----WEYIQDPRTOOTKLEBGNMHTDPDIRFPDVPVPAQYRYYPE 576
Oy 522 --NEKEMTQKDLITSDMADYQOPL-MIGTGVYTRKSGTFRPMOTDAEAG 569
Db 577 RWSFAGMRLEVLGCDWTDSPKPYETELGPTVKSSETTPPYMDADATECG 627

```

RESULT 5

```

US-08-936-135-10
: Sequence 10, Application US/08936135
: Patent No. 6054293
: GENERAL INFORMATION:
: APPLICANT: Tessier-Lavigne, Marc
: APPLICANT: He, Zhigang
: APPLICANT: Chen, Hang
: TITLE OF INVENTION: Semaphorin Receptors
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 75 DENISE DRIVE
: CITY: HILLSBOROUGH
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94010
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,135
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UC97-288-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 343-4341
: TELEFAX: (650) 343-4342
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 909 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-936-135-10

```

Query Match

```

9.4%: Score 354.5; DB: 3; Length 909;
Best Local Similarity 21.5%; Pred. No. 2.4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

```

```

Oy 41 CGHETTYDSTGTSKPNPGTYPNHTVEKTIYVK-GKRLILRLG-DLQIESQTCASDY 98
Db 28 CGGRNSMDAGYITISPGYIPDYPSNQCENWYVAPENQKIYLVNPNPFIETEKHKDKIDF 87

```

```

Oy 99 ILF-----TSSDQYGYCGSMTPKELLNTSEVTVRESSGSHISGRGLITY-----AS 149
Db 88 IETIRGDSNADLKKHCGNT-APPTIISGSGVLTKTSDARAGAGFSLRYETFKGS 146
Oy 150 SD-----HPDLITC---LERASHYLTXYSKP-----173
Db 147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPMHEILLQPLFDLEHDLQVCEG 206
Oy 174 -CPACRPAVAGDIS--GNWVQYRDTSLCKRAHAGIADLGQGVYLRKGSIRY--228
Db 207 DCKYIMLDIMGCIPIVAGPLIGCYCTKTPSKLRSTGLISLFTFMDMAADGFSARYYL 266
Oy 229 -----EGLANGVLSDGSLSDKRLFTSNGCSRSLSPEPDQIRAS 270
Db 267 IHQEPENPQCVPLGMSGRIANEQISASTFSDGR-----303
Oy 271 SSMQSVNMSGQVHNSPQOARLQDQGPMSASDSSNNHKKPRMWEIDYGEKKKITGITTT 330
Db 304 -----WTPOOSRLHGDNGMTPNLDN---KEYLOVDLRLFLMLTAIATQ 345
Oy 331 G--STOSNENFYKGFVAMNFKNNNSKMTYKGIYVNEKEVFGNSFNDPVOONNFIPIY 388
Db 346 GAISETOKGYIVKSKYLEVSTNGEDMAYYRHKNH--KIFQANDATEVYLNKLMPL 403
Oy 389 ARYVAVPQTMHQRALAEVLIQCOITQGNDSLVRKTS---QSTSVSTKKEDETIPI 445
Db 404 TRFIRPOTMILGIALRLLEFGCHVTAPCSNMLGMLSGLIADQISASTREYIMSPS 463
Oy 446 -----PSEETSGINITVAIPVLVLYVYAGMIFA---APRK 482
Db 464 AARLVSSSGWFPRNPQAPGEEW-LQYDLGTPKTVKGVIIQANGGDSITAVEARAFVR 522
Oy 483 KKKKSPYGSAAEAKTDCKQIKYEPFARHQA-EFTISYD-----521
Db 523 KFKVSYSLNGKD-----WEYIQDPRTOOTKLEBGNMHTDPDIRFPDVPVPAQYRYYPE 576
Oy 522 --NEKEMTQKDLITSDMADYQOPL-MIGTGVYTRKSGTFRPMOTDAEAG 569
Db 577 RWSFAGMRLEVLGCDWTDSPKPYETELGPTVKSSETTPPYMDADATECG 627

```

RESULT 6

```

US-08-936-135-12
: Sequence 12, Application US/08936135
: Patent No. 6054293
: GENERAL INFORMATION:
: APPLICANT: Tessier-Lavigne, Marc
: APPLICANT: He, Zhigang
: APPLICANT: Chen, Hang
: TITLE OF INVENTION: Semaphorin Receptors
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 75 DENISE DRIVE
: CITY: HILLSBOROUGH
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94010
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,135
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UC97-288-2
: TELECOMMUNICATION INFORMATION:

```

```

Oy 41 CGHETTYDSTGTSKPNPGTYPNHTVEKTIYVK-GKRLILRLG-DLQIESQTCASDY 98
Db 28 CGGRNSMDAGYITISPGYIPDYPSNQCENWYVAPENQKIYLVNPNPFIETEKHKDKIDF 87

```


OY 522 --NEKEMTKIDLITSDMADYQOPL-MIGCTVTRKSGTFPRPMDTAEAG 569
Db 577 RWSBAGIGMRLEVLGCDWTDKPTVETLGPVKSEETTPYPMDEDATECG 627

RESULT 8

US-08-936-135-16
Sequence 16, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 931 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-16

Query Match 9.4% Score 354.5; DB 3; Length 931;
Best Local Similarity 21.5%; Pred. No. 2.5e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

OY 41 CGHLVYQDSGTMTSKNYPGYPHNTVEKTIIVPK-GKRLILRLG-DLDESQTCASDY 98
Db 28 CGGRPNKSDAGYITSPGYPODYPHONCEWIVVAPEPNQKIVLNFPHFEIEKDKCKYDF 87
OY 99 LLE-----TSSDQGYPCGSMVPEKLLNTSEYTVARESSSHISGRFLTTY-----AS 149
Db 88 IEIRDGSESADLLGKRGNT-APPTIISGSVYIKFTSDYARQAGAFSLRYELFKTGS 146
OY 150 SD-----HPDLITC-----LERASHVLTETYSKF----- 173
Db 147 EDCSKNTSPNGTIESGCFPEKRYPHNLDCTITLAKPMELIILQFLPTDLHDPLOVGE 206
OY 174 -CPAGCRDVAGDIS--GNMVDGYRDTSLCKRAIHAGIADLGQISLQKRGISRY-- 228
Db 207 DCKYDMDIMGIRHYVGLKCYGCTKPSKLSSTGILSLPHHDMAVANDGSARYLL 266
OY 229 -----EGLANGVLSRDGSLDKRFLFTSNGCSLSLSTPPDQIRAS 270
Db 267 IHQPPENFOCNVPLGMSGRIANEOISASTFSDGR----- 303
OY 271 SSMQSVNESDQVHMSPGQARLODQSPMAGSDSNHMKPREWLEIDLEKKKLTIGITTT 330

Db 304 -----WTPQOSRLHGDDNGWTPNLDN-----KEYIQVDLRELTMLTALATQ 345
OY 331 G-STOSNFNFYVASFVAFNPNNNNSKMTYKGIYVNEEKVPOGNSNFRDPYONNFIPIV 388
Db 346 GAISRETKGYVYSKYLEVSTNGEDMMVYRGNKH--KIQANNDAVEYVLLNLHPL 403
OY 389 ARYRVVPTWTHORIALKVELIGCQITGQNDLSYWRKTS--OSTSVTKKEDETTIRPI 445
Db 404 TRFIRIPQVHMLIALLELFGCRVTDAPCSNMIGMLSGILADTOISASTREYLSPS 463
OY 446 -----PSEETSGINITYVAIPVLVYVAFAGGIFA-----AFRK 482
Db 464 AARLVSSRSGWFPNPNPOAPEEW-LOYDLGTPKTVKGVITIGARGGSLTAVARAVR 522
OY 483 KKKKSPYSAEAKQTKDCKQIKYPPARHQA-EFTISYD----- 521
Db 523 KFKVSYSLNGMD-----WEYIDPRTQGTFLFGSNHMYDPRDRPDPVAYRYVPE 576
OY 522 --NEKEMTKIDLITSDMADYQOPL-MIGCTVTRKSGTFPRPMDTAEAG 569
Db 577 RWSBAGIGMRLEVLGCDWTDKPTVETLGPVKSEETTPYPMDEDATECG 627

RESULT 9

US-08-936-135-18
Sequence 18, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-18

Query Match 9.3% Score 351.5; DB 3; Length 909;
Best Local Similarity 22.2%; Pred. No. 4.5e-22;
Matches 142; Conservative 99; Mismatches 255; Indels 143; Gaps 25;

OY 41 CGHLVYQDSGTMTSKNYPGYPHNTVEKTIIVPK-GKRLILRLG-DLDESQTCASDY 98
Db 28 CGGRPNKSDAGYITSPGYPODYPHONCEWIVVAPEPNQKIVLNFPHFEIEKDKCKYDF 87

STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-6

Query Match 9.0%; Score 338.5; DB 3; Length 923;
Best Local Similarity 20.8%; Pred. No. 6, 5e-21;

Matches 158; Conservative 106; Mismatches 267; Indels 227; Gaps 30;

17 RGL-----LALLIIVAPRLQAEELGCGHLVYQDSGTMTSKNPGYPHNHYCEK 70
1 RGL-----LALLIIVAPRLQAEELGCGHLVYQDSGTMTSKNPGYPHNHYCEK 70
3 RGL-----LALLIIVAPRLQAEELGCGHLVYQDSGTMTSKNPGYPHNHYCEK 70
71 TITVPGK-KRLILRLG-DLDESOTCASDYLFTSSSDQ-----YGPYCGSMVTPKELLN 124
57 LIAPEPYQRIIMFNPHFLEDDCKYDVEYIDGENEGRLMGKFCGKI-APSPVYSS 115
125 TSEYTVFESGSHISGRGFLTYA-----SSDHPDLITC-- 158
116 GPFLFIKIVSDYETHGAGFSIRYEIFKRGPCSQNTAPGVIKSPGPEKYPRNSLECTY 175
159 -----LEPASHYLTETYSKFCPAG--CRDVGADISGNMVDGYRDTSLICKAAIH 205
176 IIFAPKMESEILFEFESFDLEDQSNP--PGGMFCRYDLEI-----WDGFPE-----VGFPH 223
206 AG-IIADELGQI-----SVLQRKGISRYEGILANGVLSRQGLSDKRLF 250
224 IGRYCGQTPERIKRSSGGLSVMTYDTSALAKEGFSANYSVLQSSI--SEDEK----- 274
251 TSNCGSRSLSEEP-----DGOIRASSSMOSVNESGDQVHWSFGQARLADOGSPMASGDSN 306
275 -----CMALGMESEIHSDDOTASSOYGT-----NMSYVERSLNTPENGMTIPGDESY 322
307 NHRERLELDGKKKITTIGRTTG--STOSNENFYVKSFFYANFKNNNSKRYTKGIYNN 364
323 -----KEMIOVDLGLRRTVAVGTGAIKSKETKKYKTYKTVYDVISNGEDWISLK--EGN 376
365 EEKVFQNSNRPDYPVONNFIPIYARYVAVPOTHORIALKVELLICQITQGNDSLWVR 424
377 KATFOGNTMTDVAVLVESKPLITRFRVRIKPVSMETGISMRFEVYCGKITD-----YP 430
425 KTSOSTSVSTRKEDEITRPIPESETSTGINITTV-----AIPVLVLLVAVFAGMGIFA 478
431 CSGLGAVSGILSDQITASQADRNMMPENIRLVTSRTGALP----- 474
479 AFRKKKKKSGPYGSAEAKTDCKWKOI-----KTPPARHOSAEFTISYDN 522
475 -----PSPHY-----INEMLOVDLGDDEKIVRGVYILOGKHKREKVPFKKFIAYSN 521
523 EKEWTQALDITSDMAYQOPLMIGTGVTRKSGTFRPMOTDAEAGVSDAGHYDPOQ 582

Db 522 -----NSDWKTIKMDSKRKAQSPFGNNNDTPE 550
QY 583 RAGRHEVALPLAPEPEYATPI-----VERHYLRATFESQSGRYVPGQPGKHSLSSGCF 639
Db 551 LRTFSPSTRIRIYPERATHSGILRRELGCVEEATPAGPTTPGNCNPVHE----- 602
QY 640 SPVAVGAGQDGDYORPHSAQPADRGYDRPKAVSALATE 677
Db 603 -----CDDQANCHSGTGDD--FOLTGTGYLATE 630

RESULT 12
US-09-116-473-4
Sequence 4, Application US/09116473
Patent No. 6428965
GENERAL INFORMATION:
APPLICANT: Kolodkin, Alex
APPLICANT: Gintly, David
TITLE OF INVENTION: SEMAPHORIN RECEPTOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff
STREET: 1001 G Street, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/116,473
FILING DATE: 17-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/052,762
FILING DATE: 17-JUL-1998
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32141
REFERENCE/DOCKET NUMBER: 01107,74973
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 922 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-116-473-4
Query Match 8.8%; Score 331; DB 4; Length 922;
Best Local Similarity 24.1%; Pred. No. 3e-20;
Matches 112; Conservative 81; Mismatches 166; Indels 106; Gaps 19;

Db 176 IIFAPKSEIILFESESDLEQDSNP--PGCVFCRYDLEI-----WQCFPE----- 219
QY 206 AGIADLGGQISVLQKKSISREYGLA-----NGVLSRQGSLSDKRFLFTSNG-----CS 256
Db 220 ---VGHPIGRCCKTGRIRSSGILISMFTYDSDAIKAGFSANYSVLSSISEDKCM 276
QY 257 RSLSEFEP---DQIRASSWSQVNSGDOVHNSPGQARLQDGPSPWASGDSNNHPRP 312
Db 277 EALGMESGEIHSQIVYASSQYGT-----NMSVERSLNYPENGWTPGSDSY-----RE 324
QY 313 WLEIDGEEKKTIKIRGTG--STQSNFNFYKSFVNFKNNSKWKTYKGIYNNKVFQ 370
Db 325 WIVQVDLGLAFYAVGQGAISKETKKRYVVKYTRVDISSNGEDWTLK--EGNKATIFQ 382
QY 371 GNSNFRDPVONNFIPIVARYRVVPOTHQRIAKVELIGCOT 415
Db 383 GNTNPDVYGVFPKPLITRFAIKPASMETGISMRVYGCIT 427

RESULT 13

US-07-607-538C-4.

Sequence 4, Application US/07607538C

Patent No. 5455031

GENERAL INFORMATION:

APPLICANT: Ceriani Dr., Roberto L.

APPLICANT: Peterson Dr., Jerry A.

APPLICANT: Larocca, David J.

TITLE OF INVENTION: POLYPEPTIDE WITH 46

TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING

TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,

TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-

TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-

TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF

TITLE OF INVENTION: USE THEREOF

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: V. Amzel & Assoc.

STREET: 2055 No. 5455031th Broadway

CITY: Walnut Creek

STATE: California

COUNTRY: USA

ZIP: 94596

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/607,538C

FILING DATE: 01-NOV-1990

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Viviana Amzel

REGISTRATION NUMBER: 30,930

REFERENCE/DOCKET NUMBER: CRECC-004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 943-1931

TELEFAX: (510) 943-1189

TELEX: N.A.

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 218 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE:

US-07-607-538C-4

Query Match 8.5%; Score 321; DB 1; Length 218;
Best Local Similarity 42.6%; Pred. No. 2,1e-20;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

QY 253 NCGSRSLSEFE---PDQIRASS---SWQSVNSGDOVHNSPGQARLQDGPSPWASGDS 305
Db 58 NCGSTPLGMEKNGRIENQITASSFKKSW-----GD--YNEPRARLNAGGRVNAQAKA 110
QY 306 NNHPRFMLEIDGEEKKTIKIRGTG--STQSNFNFYKSFVNFKNNSKWKTYKGIYNN 365
Db 111 NNHPRFMLEIDGEEKKTIKIRGTG--STQSNFNFYKSFVNFKNNSKWKTYKGIYNN 365
QY 366 EKVQGSNFRDPVONNFIPIVARYRVVPOTHQRIAKVELIGCOT 414
Db 169 DKIFEGNTKTHGVKNFNPITISRTVLPKTMOSIALRLLEFGDI 217

RESULT 14

US-08-162-402B-4

Sequence 4, Application US/08162402B

Patent No. 5972337

GENERAL INFORMATION:

APPLICANT: CERIANI, ROBERTO L.

APPLICANT: PETERSON, JERRY A.

APPLICANT: LARocca, DAVID J.

TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT

TITLE OF INVENTION: GLOBULE (HMFg) ANTIGEN, FRAGMENTS & FUSION PROTEIN

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder & Poplawski

STREET: 444 South Flower St., 19th Floor

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/162,402B

FILING DATE: 03-DEC-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Amzel, Viviana

REGISTRATION NUMBER: 30,930

REFERENCE/DOCKET NUMBER: P66 38215

TELECOMMUNICATION INFORMATION:

TELEPHONE: 213-622-7700

TELEFAX: 213-489-4210

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 218 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-162-402B-4

Query Match 8.5%; Score 321; DB 2; Length 218;
Best Local Similarity 42.6%; Pred. No. 2,1e-20;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

QY 253 NCGSRSLSEFE---PDQIRASS---SWQSVNSGDOVHNSPGQARLQDGPSPWASGDS 305
Db 58 NCGSTPLGMEKNGRIENQITASSFKKSW-----GD--YNEPRARLNAGGRVNAQAKA 110
QY 306 NNHPRFMLEIDGEEKKTIKIRGTG--STQSNFNFYKSFVNFKNNSKWKTYKGIYNN 365
Db 111 NNHPRFMLEIDGEEKKTIKIRGTG--STQSNFNFYKSFVNFKNNSKWKTYKGIYNN 365

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:18:56 ; Search time 36.0911 Seconds

(without alignments)
2386.400 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDCGHLTYQDSGTMT.....STKEDETTRPIPEETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 504360

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp Vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	6.7	432	11	08R327
2	28	6.7	460	11	09D696
3	28	6.7	503	11	09D4J3
4	9	2.2	440	16	08X543
5	8	1.9	216	4	096H40
6	8	1.9	255	16	092G35
7	8	1.9	264	16	08YPN2
8	8	1.9	339	10	09XEX1
9	8	1.9	364	4	09H8G1
10	8	1.9	470	4	09B8E6
11	8	1.9	519	4	096K04
12	8	1.9	524	4	096G04
13	8	1.9	536	4	096B85
14	8	1.9	612	10	09AUC3
15	8	1.9	657	4	096DM2
16	7	1.7	100	16	09PCB6

17	7	1.7	111	10	08VX51	08vxs1 cicer arlet
18	7	1.7	159	13	09DE09	09deg9 gallus gall
19	7	1.7	159	17	08TTR5	08ttr5 methanosarc
20	7	1.7	162	9	094W38	094w38 streptococc
21	7	1.7	182	5	09UAZ2	09uaz2 caenorhabdi
22	7	1.7	183	17	09HSH4	09hsh4 halobacteri
23	7	1.7	213	16	097M44	097m44 clostridium
24	7	1.7	230	10	0947Y3	0947y3 oryza sativ
25	7	1.7	231	5	09T296	09t296 caenorhabdi
26	7	1.7	233	5	09VLE1	09vle1 drosophila
27	7	1.7	237	11	09CUV2	09cuv2 mus musculu
28	7	1.7	242	2	09KRB9	09krb9 erwinia chr
29	7	1.7	254	5	021223	021223 caenorhabdi
30	7	1.7	256	5	08SQ06	08sq06 encaphalito
31	7	1.7	260	16	08X4D9	08x4d9 escherichia
32	7	1.7	261	16	09KEX1	09kex1 bacillus ha
33	7	1.7	265	10	041069	041069 plasmu sativ
34	7	1.7	270	7	098018	098018 hylobates l
35	7	1.7	270	7	098026	098026 hylobates l
36	7	1.7	272	7	098024	098024 cercopithe
37	7	1.7	272	7	098025	098025 papio sp. (
38	7	1.7	272	10	080422	080422 oryza sativ
39	7	1.7	272	16	0980V3	0980v3 mycoplasma
40	7	1.7	273	10	040628	040628 oryza sativ
41	7	1.7	275	7	098020	098020 pan paniscu
42	7	1.7	275	7	098023	098023 cercopithe
43	7	1.7	285	16	09KAK8	09kak8 bacillus ha
44	7	1.7	286	16	08ZOH9	08zoh9 salmonella
45	7	1.7	293	16	09CKB0	09ckb0 pasteurella

ALIGNMENTS

RESULT 1
ID 08R327 PRELIMINARY; PRT; 432 AA.
AC 08R327;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to RIKEN CDNA 4631413k11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026771; AAH26771.1; -
SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.4e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SCSHISGRGFLTYASSDHPDLITCER 127
DB 54 SCSHISGRGFLTYASSDHPDLITCER 81
RESULT 2
ID 09D696 PRELIMINARY; PRT; 460 AA.
AC 09D696;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 4631413K11Rik protein.
GN 4631413K11Rik
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aitawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiba H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L. M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M. F.,
RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamita M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. F.,
RA Suzuki H., Toyooka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK014521; BAB29409.1; -
DR MGD; MGI:1913936; 463143K11R1K.
DR InterPro; IPR000859; CUB_domain.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
SQ SEQUENCE 460 AA; 50334 MW; 3CF3556F70A88B93 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 460;
Best Local Similarity 100.0%; Pred. No. 4.7e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 SCSHISGRGFLTYASSDHPDLITCLER 127
DB 82 SCSHISGRGFLTYASSDHPDLITCLER 109
|||||
ID 09D4J3 PRELIMINARY; PRT; 503 AA.
AC 09D4J3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 463143K11R1K protein.
GN 463143K11R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aitawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiba H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L. M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M. F.,
RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamita M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. F.,
RA Suzuki H., Toyooka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK014521; BAB29409.1; -
DR MGD; MGI:1913936; 463143K11R1K.
DR InterPro; IPR000859; CUB_domain.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
SQ SEQUENCE 460 AA; 50334 MW; 3CF3556F70A88B93 CRC64;
```

```
RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamita M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. F.,
RA Suzuki H., Toyooka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK014521; BAB29409.1; -
DR MGD; MGI:1913936; 463143K11R1K.
DR InterPro; IPR000859; CUB_domain.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
SQ SEQUENCE 503 AA; 54547 MW; FB212E845CA068 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 SCSHISGRGFLTYASSDHPDLITCLER 127
DB 125 SCSHISGRGFLTYASSDHPDLITCLER 152
|||||
ID 08X543 PRELIMINARY; PRT; 440 AA.
AC 08X543;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Regulator of uhpf.
GN UHPC OR 2517 OR EGS4604.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N. T., Plunkett G. III, Burland V., Mau B., Glasner J. D.,
RA Rose D. J., Mayhew G. F., Evans P. S., Gregor J., Kirkpatrick H. A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E. J., Davis N. W., Lim A., Dimlant E. T., Potamousis K.,
RA Apodaca J., Anantharaman T. S., Lin J., Yen G., Schwartz D. C.,
RA Welch R. A., Blattner F. R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RA [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005598; AAG58866.1; -
DR MGD; MGI:1913936; 463143K11R1K.
DR InterPro; IPR000849; GTP_transporter.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR00881; ZAO104; 1.
DR PROSITE; PS00942; GTP; 1.
KW Complete proteome.
SQ SEQUENCE 440 AA; 48345 MW; C3442BAC11132448 CRC64;
```

Query Match 2.2%; Score 9; DB 16; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
DB 57 ILANGVLSR 65

RESULT 5

ID 096H40 PRELIMINARY; PRT; 216 AA.
AC 096H40;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
OS Similar to DNA-binding protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LONG;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008936; AAH08936.1; -
DR InterPro: IPR001909; KRAb.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF01352; KRAb.1.
DR PROSITE: PS50805; KRAb.1.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 2.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 216 AA; 25227 MW; B5C6CF182AFECF2 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 216;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
DB 50 PDLITCLE 57

RESULT 6

ID 092G35 PRELIMINARY; PRT; 255 AA.
AC 092G35;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein RC1290.
GN RC1290.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ML157;
RX MDL=21442074; PubMed-11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Welschenbach J., Claverie J.-M.,
RA Reaout D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL: AE008676; AAL03828.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 255 AA; 29018 MW; 4E4A58667E03C9B3 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 255;
Best Local Similarity 100.0%; Pred. No. 9.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 370 ALKVELIG 377
DB 242 ALKVELIG 249

RESULT 7

ID 08YPN2 PRELIMINARY; PRT; 264 AA.
AC 08YPN2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Alr4161.
GN ALR4161.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2159585; PubMed-11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kurlitz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003595; BAB75860.1; -
DR InterPro: IPR001989; Radical_activat.
DR Pfam: PF02143; Radical_activat; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 264 AA; 29421 MW; BDF27F3428A007D0 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 264;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 357 YRVVPQT 364
DB 250 YRVVPQT 257

RESULT 8

ID 09XEX1 PRELIMINARY; PRT; 339 AA.
AC 09XEX1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Tryptophan synthase (EC 4.2.1.20) (Fragment).
OS Gracilariaria veruicosa.
OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariiales; Gracilariaceae;
OC Gracilariia.
OX NCBI_TaxID=2777;
RN [1]
RP SEQUENCE FROM N.A.
RA Lilius A.O., Ragan M.A.;
RT "Occurrence of closely spaced genes in the nuclear genome of the
agapophyte Gracilariia gracilis.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-SERINE + 1-(1NDOL-3-YL)GLYCEROL 3-PHOSPHATE
= L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SIMILARITY: TO THE TRPB FAMILY.
DR EMBL: AF121272; AAD17312.1; -
DR HSSP: P00933; 2MSY
DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000993; Trp_synth_beta.
DR Pfam: PF00291; PALP; 1.

DR PROSITE; PS00168; TRP-SYNTASE_BETA; 1.
 KM Lyase; Pyridoxal phosphate; tryptophan biosynthesis.
 FT NON_TER 339
 SQ SEQUENCE 339 AA; 37422 MW; 89ADAF872162349 CRC64;

Query Match
 Best Local Similarity 1.9%; Score 8; DB 10; Length 339;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 394 OSTSVSTR 401
 |||||
 DB 2 OSTSVSTR 9

RESULT 9
 O9H8G1 PRELIMINARY; PRT; 364 AA.

ID O9H8G1
 AC O9H8G1
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-JUN-2001 (TREMBLrel. 16, Last sequence update)
 DE CDNA FLJ13659 f1s, clone PLACE1011576, moderately similar to human
 DE kruppel related zinc finger protein (HTRF10) mRNA.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACEMENT:
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
 RA Watanabe S., Hosokawa T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.,
 RA *NEDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK033721; BAB14656.1; -
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KM Dna-binding; Metal-binding; Zinc-finger.
 KW SEQUENCE 364 AA; 42472 MW; 5043DA4C575842B0 CRC64;
 SO

Query Match
 Best Local Similarity 1.9%; Score 8; DB 4; Length 364;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 82 PDLITCLE 89

RESULT 10
 O9B2E6 PRELIMINARY; PRT; 470 AA.

ID O9B2E6
 AC O9B2E6
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE KRAB zinc finger protein HZF26 (Fragment).
 GN HZF26.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95169271; PubMed=7865130;
 RA Ahrink M., Aveskog M., Hellman L.,
 RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
 RT proteins expressed in the human monoblast cell line U-937."
 RL DNA Cell Biol. 14:125-136(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ahrink M., Aveskog M., Hellman L.,
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF325191; AK01422.1; -
 DR HSSP; P08046; 1AIG.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 10.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 9.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 11.
 KM Dna-binding; Metal-binding; Nuclear protein; Zinc-finger.
 FT NON_TER 470
 SO SEQUENCE 470 AA; 54961 MW; 997CB1C95D37E983 CRC64;

Query Match
 Best Local Similarity 1.9%; Score 8; DB 4; Length 470;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 50 PDLITCLE 57

RESULT 11
 O96M04 PRELIMINARY; PRT; 519 AA.

ID O96M04
 AC O96M04
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CDNA FLJ32933 f1s, clone TEST12007466, moderately similar to zinc
 DE finger protein 91.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS:
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Hotuta T., Hiraoaka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Watanabe M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK057495; BAB1510.1; -
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 12.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN; 8.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 12.

KW DNA-binding; zinc-finger.
SQ SEQUENCE 519 AA; 60350 MW; 7E4D17B53EC7732C CRC64;

Query Match 1.9%; Score 8; DB 4; Length 519;
Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
Db 59 PDLITCLE 66

RESULT 12

096JCA PRELIMINARY; PRT; 524 AA.
AC 096JCA; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE KRAB zinc finger protein.
GN KR19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE=21303228; PubMed=11410164;
RA Mark C., Looman C., Ahrink M., Hellman L.;
RT Molecular cloning and preliminary functional analysis of two novel
RT human KRAB zinc finger proteins, HKR18 and HKR19."
RL EMBL: AF277624; AAK61307.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf_C2H2; 12.
DR PROSITE: PS0805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 12.
KW DNA-binding; zinc-finger
SQ SEQUENCE 524 AA; 60598 MW; F1AD4929DC67D105 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
Db 62 PDLITCLE 69

RESULT 13

096BBS PRELIMINARY; PRT; 536 AA.
AC 096BBS; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 61.9 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-B-CELL;
RA Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC015765; AAH15765.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf_C2H2; 13.
DR PROSITE: PS0805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 13
KW DNA-binding; Hypothetical protein; Zinc-finger.
SQ SEQUENCE 536 AA; 61932 MW; 3DC80FAE53F8CC00 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
Db 50 PDLITCLE 57

RESULT 14

09AUC3 PRELIMINARY; PRT; 612 AA.
AC 09AUC3; 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor-like protein kinase 3.
GN PRK3.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanales; Solanum.
OX NCBI_Taxid=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VF36;
RA Kim H.-U., Cotter R., McCormick S.;
RT Arabidopsis: the tomato kinases identified in tomato, maize and
RT expression patterns during pollen tube growth."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF43040; AAK28345.1; -
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR001245; Ser_thr_pkinase.
DR Pfam: PF00560; LRR; 5.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00370; LRR; 5.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; TYRKc; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 612 AA; 68096 MW; A39B3C3751C3F27 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 612;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 KELLINTS 92
|||||
Db 31 KELLINTS 38

RESULT 15

096DW2 PRELIMINARY; PRT; 657 AA.
AC 096DW2; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to RecQ protein-like 4 (Fragment).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=LYMPH;
RC Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013277; AAH13277.1; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1 1
SQ SEQUENCE 657 AA; 73415 MW; CF51B910F560CA18 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 657;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 PGQARLQD 260
|||
Db 576 PGQARLQD 583

Search completed: May 15, 2003, 13:24:08
Job time : 40.0911 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:16:36 ; Search time 42.6531 Seconds

(without alignments)
1305.856 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDGCGHLVYDSTGWT.....STKKEDETTRIPSEETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 306882

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	100.0	539	22	AAU00630
2	418	100.0	586	22	AAU00629
3	400	95.7	487	22	AAU00628
4	116	27.8	398	23	AAE22715
5	116	27.8	398	23	AAU79459
6	28	6.7	503	21	AAI9126
7	8	1.9	101	22	AAU19467
8	8	1.9	101	23	ABP51369
9	8	1.9	101	23	ABP51461
10	8	1.9	109	22	AAH38946

11	8	1.9	117	22	AAH92967	Human protein seg
12	8	1.9	117	23	ABP51401	Human MDPF SEQ ID
13	8	1.9	126	22	AAH90976	Human immune/haema
14	8	1.9	128	22	AAH79595	Human protein SEQ
15	8	1.9	193	22	AAH40732	Human polypeptide
16	8	1.9	273	23	ABG60044	Human DTHP polype
17	8	1.9	353	20	AAH39779	CBMACD04 protein s
18	8	1.9	364	22	AAH94710	Human protein seg
19	8	1.9	493	22	ABG16286	Novel human diagno
20	7	1.7	104	22	AAH79413	Corynebacterium gl
21	7	1.7	111	22	AAO03138	Human polypeptide
22	7	1.7	115	22	ABG27684	Novel human diagno
23	7	1.7	133	22	AAO05265	Human polypeptide
24	7	1.7	142	22	AAH79377	Corynebacterium gl
25	7	1.7	143	22	AAU45897	Protonibacterium
26	7	1.7	150	20	AAH36191	Human secreted pro
27	7	1.7	175	21	AAH25732	Arabidopsis thalia
28	7	1.7	182	21	AAH25731	Arabidopsis thalia
29	7	1.7	206	22	AAH92643	C glutamicum prote
30	7	1.7	233	22	ABH67793	Drosophila melano
31	7	1.7	258	22	AAH92053	C glutamicum prote
32	7	1.7	258	22	AAH79542	Corynebacterium gl
33	7	1.7	259	23	ABH25998	Streptococcus poly
34	7	1.7	272	16	AAH74043	Rice carbonic-anhy
35	7	1.7	275	22	AAH86528	P. panthotrophus GB
36	7	1.7	297	22	ABH12597	Novel human diagno
37	7	1.7	299	22	AAH90730	C glutamicum prote
38	7	1.7	312	22	AAU14605	Novel bone marrow
39	7	1.7	316	22	AAH71721	Human olfactory re
40	7	1.7	322	23	ABH49963	Listeria monocytog
41	7	1.7	350	22	AAH93217	Human polypeptide
42	7	1.7	365	23	AAH22721	Human neuropilin-H
43	7	1.7	377	20	AAH35236	Chlamydia pneumoni
44	7	1.7	382	22	AAH79412	Corynebacterium gl
45	7	1.7	383	22	AAH86252	C. glutamicum Lp-6

ALIGNMENTS

RESULT 1	
ID	AAU00630 standard; Protein: 539 AA.
XX	
AC	AAU00630;
XX	
D7	29-AUG-2001 (first entry)
XX	
DE	Novel human protein (NHP) sequence #3.
XX	
KW	Novel human protein: NHP; CUB domain; extracellular domain; gene therapy;
KW	obesity; high blood pressure; connective tissue disorder; infertility;
KW	NHP-mediated pathway.
XX	
OS	Homo sapiens.
XX	
PN	MO200129219-A1.
XX	
PD	26-APR-2001.
XX	
PE	08-OCT-2000; 2000MO-US28798.
XX	
PR	19-OCT-1999; 99US-0160285.
PR	18-FEB-2000; 2000US-0183583.
XX	
PA	(TEXT-) LEXICON GENETICS INC.
PI	
PI	Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX	
DR	WPI: 2001-290917/30.
DR	N-PSDB; AAS00615.
XX	
PT	Novel nucleic acid encoding human CUB-domain containing protein, useful

PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -

XX Claim 5; Page 29-30; 33pp; English.

CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

XX Sequence 539 AA;

Query Match 100.0%; Score 418; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 60
DB 35 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 94
QY 61 ASDYLLFTSSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITYASSDHPD 120
DB 95 ASDYLLFTSSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITYASSDHPD 154
QY 121 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDYRDTSLCKKAHAGITADBLG 180
DB 155 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDYRDTSLCKKAHAGITADBLG 214
QY 181 GOISVLOQRKGISREYGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 215 GOISVLOQRKGISREYGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 274
QY 241 SVNBSGDVHMSPGQARLQDGPMSAGDSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 300
DB 275 SVNBSGDVHMSPGQARLQDGPMSAGDSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 334
QY 301 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVONNFIPIYARYRV 360
DB 335 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVONNFIPIYARYRV 394
QY 361 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 418
DB 395 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 452

RESULT 2

AAU00629 ID AAU00629 standard; Protein: 586 AA.

XX AAU00629;

XX 29-AUG-2001 (first entry)

DE Novel human protein (NHP) sequence #2.

KW Novel human protein: NHP; CUB domain: extracellular domain; gene therapy;

KW obesity; high blood pressure; connective tissue disorder; infertility;

XX NHP-mediated pathway.

XX Homo sapiens.

EN WO200129219-A1.

XX 26-APR-2001.

XX 08-OCT-2000; 2000WO-US28798.

XX 19-OCT-1999; 99US-0160285.

XX 18-FEB-2000; 2000US-0183583.

XX (LEXI-) LEXICON GENETICS INC.

XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX WPI; 2001-290917/30.

XX N-PSDB; NAA00614.

XX Claim 2; Page 27-28; 33pp; English.

CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

XX Sequence 586 AA;

Query Match 100.0%; Score 418; DB 22; Length 586;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 60
DB 82 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 141
QY 61 ASDYLLFTSSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITYASSDHPD 120
DB 142 ASDYLLFTSSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITYASSDHPD 201
QY 121 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDYRDTSLCKKAHAGITADBLG 180
DB 202 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDYRDTSLCKKAHAGITADBLG 261
QY 181 GOISVLOQRKGISREYGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 262 GOISVLOQRKGISREYGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 321
QY 241 SVNBSGDVHMSPGQARLQDGPMSAGDSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 300
DB 322 SVNBSGDVHMSPGQARLQDGPMSAGDSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 381
QY 301 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVONNFIPIYARYRV 360
DB 382 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVONNFIPIYARYRV 441
QY 361 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 418
DB 442 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 499

```

RESULT 3
AAU00628
ID AAU00628 standard; Protein: 487 AA.
XX
AC AAU00628;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #1.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN M0200129219-A1.
XX
PD 26-APR-2001.
XX
PE 08-OCT-2000; 2000MO-US28798.
XX
PR 19-OCT-1999; 990US-0160285.
XX
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2001-290917/30.
XX
DR N-PSDB; AAS00613.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 4; Page 26; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
XX
SO Sequence 487 AA;

Query Match 95.7%; Score 400; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 MTSNNGVTPNHTVCCKTTPVGRKRLIRLGDLDIESOTCASDYLLFTSSSDQVPC 78
DB 1 MTSKNFGITPNTVCCKTTPVGRKRLIRLGDLDIESOTCASDYLLFTSSSDQVPC 60

OY 79 GSNTPVELLLNTSEVTVRESGSHISGRFLTYASSDHPDLTTCERASHYLKTEYSK 138
DB 61 GSNTPVELLLNTSEVTVRESGSHISGRFLTYASSDHPDLTTCERASHYLKTEYSK 120

OY 139 FCRPGAGDVAGDISGNMVDYRDTSLCKAIAHAGITADELGGQISVLRKGISRYGIL 198
DB 121 FCRPGAGDVAGDISGNMVDYRDTSLCKAIAHAGITADELGGQISVLRKGISRYGIL 180

```

```

OY 199 ANGLSRDGSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSMOSVNESGDDVHNSPGOARL 258
DB 181 ANGLSRDGSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSMOSVNESGDDVHNSPGOARL 240

OY 259 QDQPSWASGDSNNHPRFMLEIDLGKKRTGIRTTGTSQSNFNFYVASFVNFKNNN 318
DB 241 QDQPSWASGDSNNHPRFMLEIDLGKKRTGIRTTGTSQSNFNFYVASFVNFKNNN 300

OY 319 SKMRTYGIYNNKRVQGSNFRDPVQNNFIPPIVARYRVVQPTWHORIALKVELIGC 378
DB 301 SKMRTYGIYNNKRVQGSNFRDPVQNNFIPPIVARYRVVQPTWHORIALKVELIGC 360

OY 379 QITGNDSLVWRKTSOSTSVTKKEDETIRPIPSEETST 418
DB 361 QITGNDSLVWRKTSOSTSVTKKEDETIRPIPSEETST 400

RESULT 4
AAE22715
ID AAE22715 standard; Protein: 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoicide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cereboprotective; tranquilliser; vitruicide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
PN M0200222815-A1.
XX
PD 21-MAR-2002.
XX
PE 12-SEP-2001; 2001MO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
XX
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX
DR WPI; 2002-393966/42.
XX
DR N-PSDB; AAD35992.
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
XX
PS Claim 3; Page 123-125; 152pp; English.
XX
CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.

```

CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, platelet anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC stroke, ulcers, immune deficiencies and immune disorders. Infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy1
CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

SO Sequence 398 AA;

Query Match 27.8%; Score 116; DB 23; Length 398;
Best Local Similarity 100.0%; Pred. No. 2e-106;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGGCHLVYDQSGMTSKNYPGTPNHTVCEKTLTPKGRKLLIRLGDIDIESQTCASD 63
DB 4 GGGCHLVYDQSGMTSKNYPGTPNHTVCEKTLTPKGRKLLIRLGDIDIESQTCASD 63

OY 64 YLLFTSSDQYGPYCGSMTVPKELLNLTSEVTVRESSGSHISGRFLLTYASSDHP 119
DB 64 YLLFTSSDQYGPYCGSMTVPKELLNLTSEVTVRESSGSHISGRFLLTYASSDHP 119

RESULT 5
AAU79459
ID AAU79459 standard; Protein: 398 AA.

AC AAU79459;

DT 15-JUL-2002 (first entry)

DE Human Neuropilin-Hy1.

KW Human: neuropilin-Hy1; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.

OS Homo sapiens.

PN WO200222780-A2.

PD 21-MAR-2002.

PF 11-SEP-2001; 2001WO-US28590.

PR 11-SEP-2000; 2000US-0659671.

PR 06-SEP-2001; 2001US-0659671.

XX (TANG/) TANG T Y.

XX Tang TX;

XX WPI. 2002-351881/38.

DR N-PSDB; ABK49565.

XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation

PS Claim 3; Page 118-120; 144pp; English.

CC The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating and
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy1.

SO Sequence 398 AA;

Query Match 27.8%; Score 116; DB 23; Length 398;
Best Local Similarity 100.0%; Pred. No. 2e-106;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGGCHLVYDQSGMTSKNYPGTPNHTVCEKTLTPKGRKLLIRLGDIDIESQTCASD 63
DB 4 GGGCHLVYDQSGMTSKNYPGTPNHTVCEKTLTPKGRKLLIRLGDIDIESQTCASD 63

OY 64 YLLFTSSDQYGPYCGSMTVPKELLNLTSEVTVRESSGSHISGRFLLTYASSDHP 119
DB 64 YLLFTSSDQYGPYCGSMTVPKELLNLTSEVTVRESSGSHISGRFLLTYASSDHP 119

RESULT 6
AAB19126
ID AAB19126 standard; Protein: 503 AA.

AC AAB19126;

DT 19-FEB-2001 (first entry)

DE Polypeptide isolated from lymph node stromal cells of f5n -/- mice.

KW Lymph node stromal cell; f5n -/- mice; inflammatory disorder;
KW immune system disorder; cancer; viral disorder; HIV infection;
KW blood vessel growth; tumour necrosis factor disorder; arthritis;
KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW cardiac failure.

OS Mus sp.

XX WO200058463-A1.
 PN
 XX 05-OCT-2000.
 PD
 XX
 XX 18-FEB-2000; 2000MO-NZ00015.
 PF
 XX 25-MAR-1999; 9905-0276268.
 PR
 XX 26-AUG-1999; 9905-0383586.
 PR
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA
 XX Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble RD;
 PI Murison JG;
 PI
 DR N-PSDB; AAA96736.
 DR
 XX WPI; 2000-664924/64.
 XX
 PT Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
 PT useful for modulating growth of blood cells, for treating inflammatory
 PT and tumor necrosis factor-mediated disorders, cancer and viral
 PT disorders
 XX
 PS Claim 1; Page 68-69; 75pp; English.
 XX
 CC The present sequence represents a polypeptide sequence which is
 CC isolated from lymph node stromal cells of fsn -/- mice. The
 CC polynucleotides and their polypeptides are useful for treating an
 CC inflammatory disorder, disorder of immune system and cancer selected
 CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
 CC viral disorder, in particular HIV infection and for modulating the
 CC growth of blood vessels. The polypeptides are useful for treating a
 CC tumor necrosis factor (TNF) mediated disorder, such as those selected
 CC from arthritis, inflammatory bowel disease and cardiac failure and a
 CC fibroblast growth factor-mediated disorder. It is also useful in assays
 CC to determine biological activity, to raise antihodops, to isolate
 CC corresponding ligands or receptors, to quantify levels of protein or
 CC cognate corresponding ligand or receptors, as antiinflammatory agents,
 CC and in compositions for the treatment of skin, connective tissue and
 CC immune system diseases. The polynucleotide is useful as marker for
 CC tissue, as a chromosome marker or tags in the identification of a
 CC genetic disorder.
 CC
 CC Sequence 503 AA:
 SQ
 Query Match 6.7%; Score 28; DB 21; Length 503;
 Best Local Similarity 100.0%; Pred. No. 6.8e-19;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 100 SGGSHISGRGFLTYASSDHPDLITCLER 127
 DB 125 SGGSHISGRGFLTYASSDHPDLITCLER 152
 RESULT 7
 AA019467
 ID AA019467 standard; Protein; 101 AA.
 AC AA019467;
 XX
 XX 04-DEC-2001 (first entry)
 DT
 XX Human diagnostic and therapeutic polypeptide (DITHP) #53.
 DE
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 XX Homo sapiens.
 OS
 XX
 PN WO200162927-A2.

PD 30-AUG-2001.
 XX
 XX 21-FEB-2001; 2001MO-US06059.
 PF
 XX 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'Sa SA, Banerjee S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
 XX
 DR WPI; 2001-502867/55.
 DR N-PSDB; AAS31038.
 XX
 PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
 PT
 PS Claim 27; Page 430; 522pp; English.
 PS
 XX The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DITHP expression. For example, (I) and
 CC (II) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DITHPs, by expressing inactive proteins
 CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
 CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of

CC antibodies against DITP and in assays to identify modulators of DITP
 CC expression and activity. The anti-DITP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITP in samples (e.g. by enzyme linked immunosorbent
 CC assay (ELISA)). AAU9415-AAU9625 represent human diagnostic and
 CC therapeutic (DITP) polypeptides of the invention.

XX
 CC
 S0 Sequence 101 AA;

Query Match 1.9%; Score 8; DB 22; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 Db 59 PDLITCLE 66

RESULT 8
 ABP51369
 ID ABP51369 standard; Protein: 101 AA.

XX
 AC ABP51369;

XX
 DT 03-SEP-2002 (first entry)

XX
 DE Human MDDT SEQ ID NO 391.

XX
 KW Human; MDDT: disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytoskeletal; anti-HIV;
 KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; angiot;
 KW neuroprotective; antineumatic; antiarthritic.

XX
 OS Homo sapiens.

XX
 PN WO200240715-A2.

XX
 PD 23-MAY-2002.

XX
 PF 06-SEP-2001; 2001WO-US27628.

XX
 PR 06-SEP-2000; 2000US-230505P.

XX
 PR 06-SEP-2000; 2000US-230514P.

XX
 PR 06-SEP-2000; 2000US-230515P.

XX
 PR 06-SEP-2000; 2000US-230517P.

XX
 PR 06-SEP-2000; 2000US-230518P.

XX
 PR 06-SEP-2000; 2000US-230519P.

XX
 PR 06-SEP-2000; 2000US-230595P.

XX
 PR 06-SEP-2000; 2000US-230597P.

XX
 PR 06-SEP-2000; 2000US-230598P.

XX
 PR 06-SEP-2000; 2000US-230599P.

XX
 PR 06-SEP-2000; 2000US-230610P.

XX
 PR 06-SEP-2000; 2000US-230855P.

XX
 PR 06-SEP-2000; 2000US-230988P.

XX
 PR 07-SEP-2000; 2000US-230989P.

XX
 PR 07-SEP-2000; 2000US-230991P.

XX
 PR 07-SEP-2000; 2000US-231163P.

XX
 PR 07-SEP-2000; 2000US-231167P.

XX
 PA (INCY-) INCYTE GENOMICS INC.

XX
 PI Jackson S., Lincoln SE., Altus CM., Dufour GE., Chalup MS., Hillman J.,
 PI Jones AL., Yu JY., Wright RJ., Gietzen D., Liu TF., Yap PE., Dahl CR;
 PI Montyama MG., Bradley DL., Romaygi SD., Harris B., Roseberry AM;
 PI Gerstlin EH., Peralta CH., David MH., Panzer SR., Flores V., Dalfio A;
 PI Marwaha R., Chen AJ., Chang SC., Au AP., Iman RR;

XX
 DR WPI: 2002-527544/56.
 DR N-PSDB: ABQ72586.

XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS

XX
 PS Claim 14; Page 541-542; 618pp; English.

XX
 S0 The invention relates to an isolated human disease detection and
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDDT.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.

XX
 S0 Sequence 101 AA;

Query Match 1.9%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 Db 59 PDLITCLE 66

RESULT 9
 ABP51461
 ID ABP51461 standard; Protein: 101 AA.

XX
 AC ABP51461;

XX
 DT 03-SEP-2002 (first entry)

XX
 DE Human MDDT SEQ ID NO 483.

XX
 KW Human; MDDT: disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytoskeletal; anti-HIV;
 KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; angiot;
 KW neuroprotective; antineumatic; antiarthritic.

XX
 OS Homo sapiens.

XX
 PN WO200240715-A2.

PD 23-MAY-2002.
 XX 06-SEP-2001; 2001WO-US27628.
 PF
 XX 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230555P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230855P.
 PR 06-SEP-2000; 2000US-230888P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 XX 07-SEP-2000; 2000US-231167P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daifo A;
 PI Maraha R, Chen AJ, Chang SC, Au AP, Imman RM;
 XX WPI: 2002-527544/56.
 DR N-PSDB; ABQ72678.
 XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS
 PT
 PS Claim 14; Page 603; 618pp; English.
 XX
 CC The invention relates to an isolated human disease detection and
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I) (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDP in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDP in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDP.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDP, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, portalosis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.
 CC
 SQ Sequence 101 AA:

Query Match 1.9%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 119 PDLITCLE 126
 |||||
 Db 59 PDLITCLE 66
 RESULT 10
 AAM38946
 ID AAM38946 standard; Protein; 109 AA.
 XX
 AC AAM38946;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2091.
 XX
 KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AA158102.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 PS Example 3; SEQ ID NO 2091; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification.
XX
SQ Sequence 109 AA;
Query Match 1.9%; Score 8; DB 22; Length 109;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 PDLITCLE 126
|||||
DB 82 PDLITCLE 89
RESULT 11
AAB92967
ID AAB92967 standard; Protein: 117 AA.
XX
AC AAB92967;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11667.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11667; 2537PP + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 117 AA;
Query Match 1.9%; Score 8; DB 22; Length 117;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 PDLITCLE 126
|||||
DB 81 PDLITCLE 88
RESULT 12
ABP51401
ID ABP51401 standard; Protein: 117 AA.
XX
AC ABP51401;
XX
DT 03-SEP-2002 (first entry)
XX
DE Human MDDF SEQ ID NO 423.
XX
KW Human; MDDF; disease detection and treatment molecule polynucleotide;
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW hepatotropic; antiinflammatory; antiproliferative; cytoskeletal; anti-HIV;
KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; antiagout;
KW neuroprotective; antineumatic; antiarthritic.
XX
OS Homo sapiens.
XX
PN WO200240715-A2.
XX
PD 23-MAY-2002.
XX
PF 06-SEP-2001; 2001WO-US27628.
XX
PR 06-SEP-2000; 2000US-230505P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230619P.
PR 06-SEP-2000; 2000US-230655P.
PR 06-SEP-2000; 2000US-230988P.
PR 06-SEP-2000; 2000US-230989P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JT, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
DR WPI: 2002-527544/56.
DR N-PSDB; ABQ72618.
XX
PT Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
PT e.g. AIDS -

XX Claim 14; Page 562; 618pp; English.
PS
XX The invention relates to an isolated human disease detection and
CC treatment (MDMT) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ7449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDMT in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDMT in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDMT.
CC (I) or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDMT, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germ-line gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.
XX
SQ Sequence 117 AA:

Query Match 1.9%; Score 8; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
Db 90 PDLITCLE 97

RESULT 13
AAM90976
ID AAM90976 standard; Protein; 126 AA.
XX
AC AAM90976;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18569.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228824.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

[illegible]

Query Match	Best Local Similarity	1.9%;	Score 8;	DB 22;	Length 126;
Matches	8;	Conservative	0;	Mismatches	0;
				Indels	Gaps
Oy	61 ASDYLLFT 68				
Db	20 ASDYLLFT 27				
RESULT 14					
AA079595	AA079595 standard; Protein: 128 AA.				
AA079595;					
06-NOV-2001	(first entry)				
Human protein SEQ ID NO 3241.					
Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; hematopoiesis; tissue growth factor; immunomodulatory; cancer; leukemia; nervous system disorder; arthritis; inflammation.					
Homo sapiens.					
WO200157190-A2.					
09-AUG-2001.					
05-FEB-2001; 2001WO-US04098.					
03-FEB-2000; 2000US-046914.					
27-APR-2000; 2000US-0560875.					
20-JUN-2000; 2000US-0598075.					
19-JUL-2000; 2000US-0620325.					
01-SEP-2000; 2000US-0654936.					
15-SEP-2000; 2000US-0663561.					
20-OCT-2000; 2000US-0693325.					
30-NOV-2000; 2000US-0728422.					
(HYSE-) HYSEQ INC.					
Tang Y ¹ , Liu C, Dermanac R ² , Asundi V, Zhou P, Xu C, Cao Y, Ma Y; Zhao Q ³ , Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW; Xue AJ, Yang Y, Wejhrman T, Goodrich R;					
WPI; 2001-476283/51.					
N-PSDB; AAK52728.					
Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy -					
Claim 20; Page 293; 6221pp; English.					

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAK60302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulatory
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM60020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 SQ Sequence 128 AA;
 Query Match 1.9%; Score 8; DB 22; Length 128;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDLITCLE 126
 |||||
 Db 101 PDLITCLE 108
 RESULT 15
 AAM40732
 ID AAM40732 standard; Protein; 193 AA.
 XX
 AC AAM40732;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5663.
 XX
 KW Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PE 26-DEC-2000; 2000MO-0534263.
 XX
 PR 21-JAN-2000; 2000US-048725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI59888.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 5663; 10078bp; English.

XX
 CC the invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAK42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening, and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 193 AA;
 Query Match 1.9%; Score 8; DB 22; Length 193;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDLITCLE 126
 |||||
 Db 100 PDLITCLE 107
 Search completed: May 15, 2003, 13:22:39
 Job time : 45.6531 secs

THIS PAGE BLANK (u8p10)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:22:46 ; Search time 22.3108 Seconds
(Without alignments)
1807.034 Million cell updates/sec

Title: US-10-003-132-2-COPY_35_452

Perfect score: 418
Sequence: 1 EELGDCGCHLVYODSGTMT.....STKDEDTTRPIPSHETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 362588 seqs, 96450795 residues

Word size : 0

Total number of hits satisfying chosen parameters: 204442

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : Published Applications, AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*\n2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB pep:*\n3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*\n4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*\n5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*\n6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*\n7: /cgn2_6/ptodata/2/pubpaa/PCY05_PUBCOMB pep:*\n8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*\n9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*\n10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*\n11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*\n12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*\n13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*\n14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	100.0	421	9	US-09-759-1308-76
2	418	100.0	681	9	US-09-759-1308-75
3	28	6.7	503	9	US-10-003-132-4
4	28	6.7	503	10	US-09-823-038A-51
5	19	4.5	458	9	US-10-003-132-6
6	7	1.7	206	9	US-09-738-626-6397
7	7	1.7	258	9	US-09-738-626-5807
8	7	1.7	299	9	US-09-738-626-4484
9	7	1.7	383	10	US-09-909-849-2
10	7	1.7	385	9	US-09-712-363-148
11	7	1.7	449	9	US-09-910-186A-14
12	7	1.7	653	9	US-10-060-830-114
13	6	1.4	100	10	US-09-864-761-47749
14	6	1.4	102	10	US-09-955-866-20
15	6	1.4	108	9	US-09-949-510-1
16	6	1.4	108	9	US-09-910-150-21
17	6	1.4	110	10	US-09-815-242-10553
18	6	1.4	111	10	US-09-925-300-1524
19	6	1.4	112	10	US-09-910-150-19

20	6	1.4	116	10	US-09-800-729-183	Sequence 183, App
21	6	1.4	118	10	US-09-910-150-23	Sequence 23, App1
22	6	1.4	119	10	US-09-867-550-410	Sequence 410, App
23	6	1.4	123	1	US-08-899-112-10	Sequence 10, App1
24	6	1.4	123	8	US-08-424-350B-64	Sequence 64, App1
25	6	1.4	123	10	US-09-771-161A-171	Sequence 171, App
26	6	1.4	125	9	US-09-764-668-911	Sequence 911, App
27	6	1.4	135	10	US-09-925-301-1529	Sequence 1529, App
28	6	1.4	147	9	US-09-809-391-503	Sequence 503, App
29	6	1.4	147	10	US-09-864-761-48692	Sequence 48692, A
30	6	1.4	151	9	US-09-776-724A-262	Sequence 262, App
31	6	1.4	161	10	US-09-925-301-929	Sequence 929, App
32	6	1.4	175	9	US-10-108-605-97	Sequence 97, App1
33	6	1.4	176	12	US-10-062-254-230	Sequence 230, App
34	6	1.4	178	10	US-09-737-178-132	Sequence 132, App
35	6	1.4	178	10	US-09-864-761-33694	Sequence 33694, A
36	6	1.4	178	10	US-09-864-761-46479	Sequence 46479, A
37	6	1.4	183	9	US-10-116-255-27	Sequence 27, App1
38	6	1.4	183	10	US-09-910-174A-4	Sequence 4, App1
39	6	1.4	191	9	US-10-260-877-56	Sequence 56, App1
40	6	1.4	195	10	US-09-764-864-854	Sequence 854, App
41	6	1.4	201	9	US-09-924-340-56	Sequence 56, App1
42	6	1.4	201	9	US-09-992-600A-56	Sequence 56, App1
43	6	1.4	201	10	US-09-873-438-2	Sequence 2, App1
44	6	1.4	202	9	US-10-171-311-202	Sequence 202, App1
45	6	1.4	202	10	US-09-873-438-4	Sequence 4, App1

ALIGNMENTS

RESULT 1
US-09-759-1308-76
Sequence 76, Application US/09759130B
Publication NO. US2003002279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kirst, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Myers, Paul S
APPLICANT: Leiby, Kevin R
APPLICANT: Wrighton, Nicolas
APPLICANT: Goodearl, Andrew
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE,
FILE REFERENCE: MPIOO-5350ANIM
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707

```
;; PRIOR FILING DATE: 1999-10-19
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 76
;; LENGTH: 421
;; TYPE: PR
;; ORGANISM: Homo sapiens
US-09-759-1308-76

Query Match
Best Local Similarity 100.0%; Score 418; DB 9; Length 421;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BELGDCGHLVYQDSGTMSTKNPNTVCCKITVPKGRLLIRLGLDIESQTC 60
DB 1 BELGDCGHLVYQDSGTMSTKNPNTVCCKITVPKGRLLIRLGLDIESQTC 60
QY 61 ASDYLFTSSSDQYGPYCGSMTPKELLNTSEVTVRESGSHSGRFLTYASSDHPD 120
DB 61 ASDYLFTSSSDQYGPYCGSMTPKELLNTSEVTVRESGSHSGRFLTYASSDHPD 120
QY 121 LITCLERASHLYKTEYSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAIIHAGITADELG 180
DB 121 LITCLERASHLYKTEYSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAIIHAGITADELG 180
QY 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
QY 241 SVNESGDVHMSPGARLQDGPMSAGDSSNNHPRMLELIDLEKKKITGIRTTGSGTQ 300
DB 241 SVNESGDVHMSPGARLQDGPMSAGDSSNNHPRMLELIDLEKKKITGIRTTGSGTQ 300
QY 301 SNFNFYVKSFWANFKNNSKMKTYKGIYVNEEKYFQGSNFRDPVQNNFIPPIVARYRV 360
DB 301 SNFNFYVKSFWANFKNNSKMKTYKGIYVNEEKYFQGSNFRDPVQNNFIPPIVARYRV 360
QY 361 VPQTHQRIALKVELIGQITQGNDSLWKRKTSOSTSVTKKEDETITRPIPSEETST 418
DB 361 VPQTHQRIALKVELIGQITQGNDSLWKRKTSOSTSVTKKEDETITRPIPSEETST 418

RESULT 2
US-09-759-1308-75
;; Sequence 75, Application US/09759130B
;; Publication No. US20030022279A1
;; GENERAL INFORMATION:
;; APPLICANT: Millennium Pharmaceuticals, Inc.
;; APPLICANT: McCarthy, Sean A
;; APPLICANT: Fraser, Christopher C
;; APPLICANT: Sharp, John D
;; APPLICANT: Barnes, Thomas S
;; APPLICANT: Kirst, Susan J
;; APPLICANT: Mackay, Charles R
;; APPLICANT: Myers, Paul S
;; APPLICANT: Leiby, Kevin R
;; APPLICANT: Wrighton, Nicolas
;; APPLICANT: Goodearl, Andrew
;; APPLICANT: Holtzman, Douglas A
;; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
;; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
;; TITLE OF INVENTION: USES.
;; FILE REFERENCE: MP100-5350NMIM
;; CURRENT APPLICATION NUMBER: US/09/759,130B
;; CURRENT FILING DATE: 2002-09-16
;; PRIOR APPLICATION NUMBER: US 09/479,249
;; PRIOR FILING DATE: 2000-01-07
;; PRIOR APPLICATION NUMBER: US 09/559,497
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 09/578,063
;; PRIOR FILING DATE: 2000-05-24
;; PRIOR APPLICATION NUMBER: US 09/333,159
;; PRIOR FILING DATE: 1999-06-14
```

```
;; PRIOR APPLICATION NUMBER: US 09/596,194
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 09/342,364
;; PRIOR FILING DATE: 1999-06-29
;; PRIOR APPLICATION NUMBER: US 09/608,452
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/393,996
;; PRIOR FILING DATE: 1999-09-10
;; PRIOR APPLICATION NUMBER: US 09/602,871
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: US 09/420,707
;; PRIOR FILING DATE: 1999-10-19
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 75
;; LENGTH: 681
;; TYPE: PR
;; ORGANISM: Homo sapiens
US-09-759-1308-75

Query Match
Best Local Similarity 100.0%; Score 418; DB 9; Length 681;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BELGDCGHLVYQDSGTMSTKNPNTVCCKITVPKGRLLIRLGLDIESQTC 60
DB 1 BELGDCGHLVYQDSGTMSTKNPNTVCCKITVPKGRLLIRLGLDIESQTC 60
QY 61 ASDYLFTSSSDQYGPYCGSMTPKELLNTSEVTVRESGSHSGRFLTYASSDHPD 120
DB 61 ASDYLFTSSSDQYGPYCGSMTPKELLNTSEVTVRESGSHSGRFLTYASSDHPD 120
QY 121 LITCLERASHLYKTEYSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAIIHAGITADELG 180
DB 121 LITCLERASHLYKTEYSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAIIHAGITADELG 180
QY 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
QY 241 SVNESGDVHMSPGARLQDGPMSAGDSSNNHPRMLELIDLEKKKITGIRTTGSGTQ 300
DB 241 SVNESGDVHMSPGARLQDGPMSAGDSSNNHPRMLELIDLEKKKITGIRTTGSGTQ 300
QY 301 SNFNFYVKSFWANFKNNSKMKTYKGIYVNEEKYFQGSNFRDPVQNNFIPPIVARYRV 360
DB 301 SNFNFYVKSFWANFKNNSKMKTYKGIYVNEEKYFQGSNFRDPVQNNFIPPIVARYRV 360
QY 361 VPQTHQRIALKVELIGQITQGNDSLWKRKTSOSTSVTKKEDETITRPIPSEETST 418
DB 361 VPQTHQRIALKVELIGQITQGNDSLWKRKTSOSTSVTKKEDETITRPIPSEETST 418

RESULT 3
US-10-003-132-4
;; Sequence 4, Application US/10003132
;; Publication No. US20020192750A1
;; GENERAL INFORMATION:
;; APPLICANT: Fox, Brian A.
;; APPLICANT: Gao, Zeren
;; APPLICANT: Shoemaker, Kimberly E.
;; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
;; FILE REFERENCE: 00-62
;; CURRENT APPLICATION NUMBER: US/10/003,132
;; CURRENT FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US 60/249,004
;; PRIOR FILING DATE: 2000-11-15
;; NUMBER OF SEQ ID NOS: 19
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 4
;; LENGTH: 503
;; TYPE: PR
;; ORGANISM: Mus musculus
```

US-10-003-132-4

Query Match 6.7%; Score 28; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SCSHISGRGFLTYASSDHPDLITCLER 127
DB 125 SCSHISGRGFLTYASSDHPDLITCLER 152

RESULT 4
US-09-823-038A-51
Sequence 51, Application US/09823038A
Patent No. US20020058335A1

GENERAL INFORMATION:
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Onrust, Rene
APPLICANT: Kumbie, Anand
APPLICANT: Murison, Greg
TITLE OF INVENTION: Compositions Isolated From Stromal Cells
FILE REFERENCE: 11000.1037C3
CURRENT APPLICATION NUMBER: US/09/823,038A
CURRENT FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 51
LENGTH: 503
TYPE: PRF
ORGANISM: Mouse
US-09-823-038A-51

Query Match 6.7%; Score 28; DB 10; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SCSHISGRGFLTYASSDHPDLITCLER 127
DB 125 SCSHISGRGFLTYASSDHPDLITCLER 152

RESULT 5
US-10-003-132-6
Sequence 6, Application US/10003132
Publication No. US20020192750A1

GENERAL INFORMATION:
APPLICANT: Fox, Brian A.
APPLICANT: Gao, Zeren
APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUTROFILIN HOMOLOG ZCUB5
FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003,132
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/249,004
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 458
TYPE: PRF
ORGANISM: Mus musculus
US-10-003-132-6

Query Match 4.5%; Score 19; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 DGYRDTSLCKAIIHAGII 175
DB 137 DGYRDTSLCKAIIHAGII 155

RESULT 6
US-09-738-626-6397
Sequence 6397, Application US/09738626
Publication No. US20020197605A1

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6397
LENGTH: 206
TYPE: PRF
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6397

Query Match 1.7%; Score 7; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 TYRFEFG 101
DB 75 TYRFEFG 81

RESULT 7
US-09-738-626-5807
Sequence 5807, Application US/09738626
Publication No. US20020197605A1

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5807
LENGTH: 258

;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5807

Query Match 1.7%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 WASGDS 271
|||||
DB 240 WASGDS 246

RESULT 8
US-09-738-626-4484
; Sequence 4484, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SEMOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4484
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4484

Query Match 1.7%; Score 7; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 TYPKELL 88
|||||
DB 16 TYPKELL 22

RESULT 9
US-09-909-849-2
; Sequence 2, Application US/09909849
; Patent No. US20020106754A1
; GENERAL INFORMATION:
; APPLICANT: Tauch, Andreas
; TITLE OF INVENTION: Nucleotide Sequences Which Code for the *atr* Gene
; FILE REFERENCE: 032301 WD 173
; CURRENT APPLICATION NUMBER: US/09/909, 849
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-909-849-2

Query Match 1.7%; Score 7; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 SWAGDS 270
|||||
DB 376 SWAGDS 382

RESULT 10
US-09-712-363-148
; Sequence 148, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712, 363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179, 531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117, 844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118, 206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126, 593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134, 093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134, 092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165, 124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165, 086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-148

Query Match 1.7%; Score 7; DB 9; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LILRLGD 52
|||||
DB 280 LILRLGD 286

RESULT 11
US-09-910-186A-14
; Sequence 14, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Materiel Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626-A 067252.0107
; CURRENT APPLICATION NUMBER: US/09/910, 186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611, 419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/133, 865

;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,866
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,867
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,868
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,869
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,873
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 08/123,975
;; PRIOR FILING DATE: 1993-09-21
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 14
;; LENGTH: 449
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic Construct
US-09-910-186A-14

Query Match 1.7%; Score 7; DB 9; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 312 MNEKNN 318
|||||||
Db 396 MNEKNN 402

RESULT 12
US-10-060-830-1114
;; Sequence 1114, Application US/10060830
;; Publication No. US20030032154A1
;; GENERAL INFORMATION:
;; APPLICANT: Gu, Yizhong
;; APPLICANT: Nguyen, Cung-Tuong
;; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
;; FILE REFERENCE: PB0169
;; CURRENT APPLICATION NUMBER: US/10/060,830
;; CURRENT FILING DATE: 2002-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/325,062
;; PRIOR FILING DATE: 2001-09-25
;; NUMBER OF SEQ ID NOS: 1123
;; SOFTWARE: Aeomica Sequence Listing Engine
;; SEQ ID NO 1114
;; LENGTH: 653
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-060-830-1114

Query Match 1.7%; Score 7; DB 9; Length 653;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 179 LGGQISV 185
|||||||

Db 128 LGGQISV 134

RESULT 13
US-09-864-761-47749
;; Sequence 47749, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;; FILE REFERENCE: Aeomica-x-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 47749
;; LENGTH: 100
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL109809.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.88
;; OTHER INFORMATION: EST HUMAN HIT: AW406955.1, EVALUATE 1.00e-01
;; OTHER INFORMATION: SWISSPROT HIT: P01871, EVALUATE 7.00e-03
US-09-864-761-47749

Query Match 1.4%; Score 6; DB 10; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 LTVASS 116
|||||||
Db 11 LTVASS 16

RESULT 14

US-09-955-866-20
 : Sequence 20, Application US/09955866
 : Patent No. US20020107363A1
 : GENERAL INFORMATION:
 : APPLICANT: Fox, Michael
 : APPLICANT: Sullivan, John K.
 : APPLICANT: Holst, Paige
 : APPLICANT: Yoshinaga, Steven Kiyoshi
 : TITLE OF INVENTION: B7-Like Polypeptides and Uses Thereof
 : FILE REFERENCE: 00,759-A
 : CURRENT APPLICATION NUMBER: US/09/955,866
 : CURRENT FILING DATE: 2001-09-19
 : PRIOR APPLICATION NUMBER: 60/233,867
 : PRIOR FILING DATE: 2000-09-20
 : NUMBER OF SEQ ID NOS: 30
 : SOFTWARE: Patentln Ver. 2.0
 : SEQ ID NO 20
 : LENGTH: 102
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-09-955-866-20

Query Match

1.43; Score 6; DB 10; Length 102;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 TPKEL 87
 |||||
 Db 6 TPKEL 11

RESULT 15

US-09-949-510-1
 : Sequence 1, Application US/09949510
 : Publication No. US20030077840A1
 : GENERAL INFORMATION:
 : APPLICANT: Chait et al.
 : TITLE OF INVENTION: METHOD FOR THE COMPARATIVE QUANTITATIVE ANALYSIS OF
 : TITLE OF INVENTION: PROTEINS AND OTHER BIOLOGICAL MATERIAL BY ISOTOPIC
 : TITLE OF INVENTION: LABELING AND MASS SPECTROSCOPY
 : FILE REFERENCE: Seq. List 1-6
 : CURRENT APPLICATION NUMBER: US/09/949,510
 : CURRENT FILING DATE: 2001-09-06
 : NUMBER OF SEQ ID NOS: 5
 : SOFTWARE: Patentln Ver. 2.1
 : SEQ ID NO 1
 : LENGTH: 108
 : TYPE: PRT
 : ORGANISM: Human
 US-09-949-510-1

Query Match

1.43; Score 6; DB 9; Length 108;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 ISRYEG 196
 |||||

Db 47 ISRYEG 52

Search completed: May 15, 2003, 13:26:35
 Job time : 25.3108 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:21:41 ; Search time 36.7473 Seconds
(without alignments)
1093.529 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDGCGHLVYDSCGTMT.....STKKDETTTRIPSEST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 223049

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.2	440	1 RGEUC	regulatory protein
2	9	2.2	440	2 D91204	regulator of uhpr
3	9	2.2	440	2 F86050	regulator of uhpr
4	8	1.9	154	2 D39384	finger protein HPR
5	8	1.9	163	2 B39384	finger protein HPR
6	8	1.9	195	2 E39384	finger protein HPR
7	8	1.9	196	2 A39384	finger protein HPR
8	8	1.9	255	2 B97861	hypothetical prote
9	8	1.9	264	2 AB2326	hypothetical prote
10	8	1.9	595	2 G02075	transcription repr
11	8	1.9	624	2 S50650	hypothetical prote
12	7	1.7	100	2 AB2629	hypothetical prote
13	7	1.7	117	2 T49511	hypothetical prote
14	7	1.7	167	2 T33602	hypothetical prote
15	7	1.7	182	2 T34009	hypothetical prote
16	7	1.7	183	2 D84183	hypothetical prote
17	7	1.7	213	2 E96943	GMP synthase subu
18	7	1.7	248	2 D69713	similar to hypot
19	7	1.7	254	2 T23323	vegetative growth
20	7	1.7	260	2 F85630	hypothetical prote
21	7	1.7	261	2 C83756	hypothetical prote
22	7	1.7	265	2 S20988	transcription regu
23	7	1.7	272	2 B90544	lectin - garden pe
24	7	1.7	273	2 T03254	hypothetical prote
25	7	1.7	285	2 S27365	probable carbonate
26	7	1.7	285	2 G83934	lectin precursor -
27	7	1.7	293	2 G64050	hypothetical prote
28	7	1.7	301	2 T19399	N-acetylneuraminat
29	7	1.7	306	2 D70601	hypothetical prote
					UTP-glucose-1-phos

30	7	1.7	306	2 T45453	UTP-glucose-1-phos
31	7	1.7	322	2 AC1759	conserved hypothet
32	7	1.7	332	2 A11383	conserved hypothet
33	7	1.7	339	2 T37727	probable nuclear p
34	7	1.7	356	1 S00139	photosynthetic rea
35	7	1.7	357	2 C69223	anion permease - M
36	7	1.7	359	2 T21373	hypothetical prote
37	7	1.7	385	2 S70984	recf protein - Myc
38	7	1.7	394	2 A70842	probable anti prot
39	7	1.7	397	2 F85681	unknown protein en
40	7	1.7	399	2 T01035	hypothetical prote
41	7	1.7	400	2 F81419	probable efflux pr
42	7	1.7	405	2 C71462	hypothetical prote
43	7	1.7	414	2 F29826	hypothetical prote
44	7	1.7	444	2 S34280	polygalacturonase
45	7	1.7	445	2 S34266	polygalacturonase

ALIGNMENTS

RESULT 1

RGEUC Regulatory protein uhpc - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 30-Jun-1991 #sequence_revision 14-Nov-1997 #extl_change 01-Mar-2002

C:Accession: D65168; G41853; C26925; S30078

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; M01D:97426617; PMID:9278503

A:Accession: D65168

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-440 <BLAT>

A:Cross-references: GB:AE000444; GB:U00096; NID:q2367258; PIDN:AC76690.1; PID:q23672

A:Experimental source: strain K-12, substrain MG1655

J. Bacteriol. 174, 2754-2762, 1992

A:Title: Structure and function of the uhpc genes for the sugar phosphate transport sy

A:Reference number: A41853; M01D:9224930; PMID:1569007

A:Accession: G41853

A:Molecule type: DNA

A:Residues: 2-440 <RSL>

A:Cross-references: GB:M89479; NID:q148116; PIDN:AA24726.1; PID:q148119

A:Note: this is a revision to the sequence from reference A30395

R:Frledrich, M.J.; Kadner, R.J.

J. Bacteriol. 169, 3556-3563, 1987

A:Title: Nucleotide sequence of the uhpc region of Escherichia coli.

A:Reference number: A30395; M01D:87279903; PMID:3301805

A:Accession: C26925

A:Molecule type: DNA

A:Residues: 'M', '87', 'VCRAIAE', '96', 'L', '98', 'P', '100', 'L', '102', 'FPG', '139-141', 'SVNGLVFTYRA', '1

A:Note: This sequence has been revised in reference A41853

C:Comment: This is one of the proteins involved in the expression of uhpr, a gene for

NAD transmembrane orientation.

C:Genetics:

A:Gene: uhpc

A:Map position: 82 min

C:Superfamily: hexose phosphate transport protein uhpr

C:Keywords: membrane protein; sugar phosphate transport system

Query Match 2.2% Score 9; DB 1; Length 440;

Best local similarity 100.0%; Pred. No. 0.62;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 197 ILANGVLSR 205

Db 57 ILANGVLSR 65

RESULT 2

D91204
regulator of uhpt [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D91204
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A96629; MUID:21156231; PMID:11258796
A:Accession: D91204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <HAV>
A:Cross-references: GB:BA000007; PIDN:BA838027.1; PID:G13364079; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECA4604
C:Superfamily: hexose phosphate transport protein uhpt

Query Match 2.2%; Score 9; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
|||||
DB 57 ILANGVLSR 65

RESULT 3
F86050
regulator of uhpt [imported] - Escherichia coli (strain O157:H7, substrain EDJ933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F86050
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Mature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F86050
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <STO>
A:Cross-references: GB:AE005174; NID:G12518499; PIDN:AAG58866.1; GSPDB:GN00145; UWGP:Z51
A:Experimental source: strain O157:H7, substrain EDJ933
C:Genetics:
A:Gene: uhpc
C:Superfamily: hexose phosphate transport protein uhpt

Query Match 2.2%; Score 9; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
|||||
DB 57 ILANGVLSR 65

RESULT 4
D39384
finger protein HTF6 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
C:Accession: D39384
R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Knieppel-associated box domain defines a subfamily
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: D39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <BEL>

A:Cross-references: GB:M61869; NID:G184449; PID:G1844450
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
DB 8 PDLITCLE 15

RESULT 5
B39384
finger protein HFP9 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 08-Dec-2000
C:Accession: B39384
R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Knieppel-associated box domain defines a subfam
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: B39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <AAA>
A:Cross-references: GB:M61867
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
DB 18 PDLITCLE 25

RESULT 6
E39384
finger protein HFP9 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 30-Jan-1993 #text_change 08-Dec-2000
C:Accession: E39384
R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Knieppel-associated box domain defines a subfam
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: E39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <BEL>
A:Cross-references: GB:M61869
C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||
DB 50 PDLITCLE 57

RESULT 7
A39384
finger protein HFP4 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
C:Accession: A39384
R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Krueppel-associated box domain defines a subfamily
A:Reference number: A39384; MID:91219421; PMID:2023909
A:Accession: A39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <EBL>
A:Cross-references: GB:M61866; MID:9454818; PIDN:AAA52689.1; PID:g184336
C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCIE 126
|||||||
DB 50 PDLITCIE 57

RESULT 8

hypothetical protein RC1290 [Imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: B97861
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ri-
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MID:21442074; PMID:11557893
A:Accession: B97861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-235 <RUR>
A:Cross-references: GB:AE006914; PIDN:AAL03828.1; PID:g15620429; GSPDB:GN00173
C:Genetics:
A:Gene: RC1290

Query Match 1.9%; Score 8; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 ALKVELIG 377
|||||||
DB 242 ALKVELIG 249

RESULT 9

AB2326
hypothetical protein alr4161 [Imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AB2326
R:Kankaki, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana-
A:Reference number: AB1807; MID:21595285; PMID:11759840
A:Accession: AB2326
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <RUR>
A:Cross-references: GB:BA000019; PIDN:BA075860.1; PID:g17133296; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4161

Query Match 1.9%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 357 YVRVPOQT 364
|||||||

DB 250 YVRVPOQT 257

RESULT 10

transcription repressor zinc finger protein 85 - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C:Accession: G02075
R:Poncellet, D.A.
submitted to the EMBL Data Library, September 1995
A:Reference number: G09169
A:Accession: G02075
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-595 <PON>
A:Cross-references: EMBL:U5376; MID:g1017721; PIDN:AAA9179.1; PID:g1017722
C:Genetics:
A:Gene: GDB:ZNF85
A:Cross-references: GDB:132279
A:Map position: 19p12-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 1.9%; Score 8; DB 2; Length 595;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCIE 126
|||||||
DB 50 PDLITCIE 57

RESULT 11

S50650
hypothetical protein YER147C - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C:Accession: S50650
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and 1amb
A:Reference number: S50430
A:Accession: S50650
A:Molecule type: DNA
A:Residues: 1-624 <DIE>
A:Cross-references: EMBL:U18917; MID:g603377; PID:g603387; GSPDB:GN00005; MIPS:YER147
C:Genetics:
A:Gene: SGD:SCC4; MIPS:YER147C
A:Cross-references: SGD:S0000949
A:Map position: 5R
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YER147C

Query Match 1.9%; Score 8; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 368 RIALKVEL 375
|||||||
DB 285 RIALKVEL 292

RESULT 12

AB2629
hypothetical protein XF1865 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2629
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A5328 below
A:Accession: AB2629

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-100 <SIV>
A:Cross-references: GB:AE004007; GB:AE003849; NID:G9106944; PIDN:AFR84671.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Agencio, M.; Alvarenga, R.; A
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincan, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XFI865

Query Match 1.7%; Score 7; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 LGDDIE 56
|||||
Db 55 LGDDIE 61

RESULT 13
T49511
hypothetical protein B14D6.630 [Imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49511
R:Schulte, U.; Aign, V.; Hohseil, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49511
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-117 <SCH>
A:Cross-references: EMBL:ALJ56173; GSPDB:GN00116; NCSP:B14D6.630
A:Experimental source: BAC clone B14D6; strain OR74A
C:Genetics:
A:Gene: NCSP:B14D6.630
A:Map position: 6
C:Superfamily: Neurospora crassa hypothetical protein B14D6.630

Query Match 1.7%; Score 7; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 GHLVYQ 14
|||||
Db 4 GHLVYQ 10

RESULT 14
T33602
hypothetical protein E02H9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33602
R:Kellen, J.; Kramer, J.; Hawkins, M.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid E02H9.
A:Reference number: 221375
A:Accession: T33602
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA

A:Residues: 1-167 <KEI>
A:Cross-references: EMBL:AF099915; PIDN:AAC68768.1; GSPDB:GN00021; CESP:E02H9.2
A:Experimental source: strain Bristol N2; clone E02H9
C:Genetics:
A:Gene: CESP:E02H9.2
A:Map position: 3
A:Insertions: 89/2

Query Match 1.7%; Score 7; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 RKGISRY 194
|||||
Db 55 RKGISRY 61

RESULT 15
T34009
hypothetical protein Y49G5A.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34009
R:Becker, M.; Tin-Wollam, A.M.; Yeakum, M.
submitted to the EMBL Data Library, February 1999
A:Description: The sequence of C. elegans cosmid Y49G5A.
A:Reference number: 221457
A:Accession: T34009
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-182 <BEC>
A:Cross-references: EMBL:AF125968; PIDN:AAD14760.1; GSPDB:GN00023; CESP:Y49G5A.1
A:Experimental source: strain Bristol N2; clone Y49G5A
C:Genetics:
A:Gene: CESP:Y49G5A.1
A:Map position: 5
A:Insertions: 44/3; 82/1; 102/3

Query Match 1.7%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 SKFCPAG 143
|||||
Db 162 SKFCPAG 168

Search completed: May 15, 2003, 13:25:55
Job time : 39.7473 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:20:51 ; Search time 24.9356 Seconds
(without alignments)
493.221 Million cell updates/sec.

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 BELGDGCGHLVYQDSGTWT.....STKKEDEITRPISERTST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 segs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 58762

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgnt2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgnt2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgnt2_6/ptodata/1/1aa/5A.COMB.pep:*
- 4: /cgnt2_6/ptodata/1/1aa/5B.COMB.pep:*
- 5: /cgnt2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
- 6: /cgnt2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	1.7	115	1 US-08-111-939-22	Sequence 22, Appl
2	7	1.7	159	2 US-08-162-402B-16	Sequence 16, Appl
3	7	1.7	218	1 US-07-607-538C-4	Sequence 4, Appl
4	7	1.7	218	2 US-08-162-402B-4	Sequence 4, Appl
5	7	1.7	272	2 US-08-492-027A-4	Sequence 4, Appl
6	6	1.4	108	4 US-09-304-799-1	Sequence 1, Appl
7	6	1.4	109	1 US-08-111-939-21	Sequence 21, Appl
8	6	1.4	109	1 US-08-111-939-25	Sequence 25, Appl
9	6	1.4	111	1 US-08-111-939-23	Sequence 23, Appl
10	6	1.4	111	4 US-09-134-001C-4964	Sequence 4964, Ap
11	6	1.4	123	2 US-08-626-685A-10	Sequence 10, Appl
12	6	1.4	123	4 US-08-469-260A-64	Sequence 64, Appl
13	6	1.4	146	1 US-08-054-480-2	Sequence 2, Appl
14	6	1.4	147	4 US-09-149-476-503	Sequence 503, App
15	6	1.4	151	2 US-08-563-148E-6	Sequence 6, Appl
16	6	1.4	151	4 US-09-194-139-1	Sequence 1, Appl
17	6	1.4	157	2 US-08-162-402B-13	Sequence 13, Appl
18	6	1.4	160	2 US-08-162-402B-14	Sequence 14, Appl
19	6	1.4	160	2 US-08-162-402B-15	Sequence 15, Appl
20	6	1.4	176	4 US-09-393-627B-17	Sequence 17, Appl
21	6	1.4	176	4 US-09-393-627B-18	Sequence 18, Appl
22	6	1.4	188	2 US-08-531-525-46	Sequence 46, Appl
23	6	1.4	188	2 US-08-718-270A-46	Sequence 46, Appl
24	6	1.4	201	2 US-08-726-228-2	Sequence 2, Appl
25	6	1.4	201	4 US-08-870-815-2	Sequence 2, Appl
26	6	1.4	201	4 US-08-949-004-2	Sequence 2, Appl
27	6	1.4	202	1 US-08-274-318-2	Sequence 2, Appl

28	6	1.4	202	2 US-08-463-081B-2	Sequence 2, Appl
29	6	1.4	202	2 US-08-461-379A-2	Sequence 2, Appl
30	6	1.4	202	2 US-08-462-390B-2	Sequence 2, Appl
31	6	1.4	202	2 US-08-754-108-2	Sequence 2, Appl
32	6	1.4	202	3 US-08-870-815-4	Sequence 4, Appl
33	6	1.4	202	3 US-08-463-074B-2	Sequence 2, Appl
34	6	1.4	202	3 US-08-465-585C-2	Sequence 2, Appl
35	6	1.4	202	3 US-08-652-446-2	Sequence 2, Appl
36	6	1.4	202	4 US-08-949-004-4	Sequence 4, Appl
37	6	1.4	208	4 US-08-961-083-214	Sequence 214, App
38	6	1.4	209	2 US-08-248-839C-58	Sequence 58, Appl
39	6	1.4	214	4 US-08-861-774E-48	Sequence 48, Appl
40	6	1.4	214	4 US-07-607-538C-5	Sequence 5, Appl
41	6	1.4	218	2 US-08-162-402B-5	Sequence 5, Appl
42	6	1.4	219	4 US-08-934-627B-6	Sequence 6, Appl
43	6	1.4	221	4 US-09-198-956-4	Sequence 4, Appl
44	6	1.4	221	4 US-09-670-141-4	Sequence 4, Appl
45	6	1.4	227	4 US-08-871-572B-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-111-939-22
; Sequence 22, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shunji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/111,939
; FILING DATE: 26-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324033/92
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 230029/92
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1321-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-111-939-22

Query Match 1.7%; Score 7; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 351 PRIVARY 357
|||||
Db 88 PRIVARY 94

RESULT 2
US-08-162-402B-16
Sequence 16, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALITON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-16

Query Match 1.7%; Score 7; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 351 PRIVARY 357
|||||
Db 129 PRIVARY 135

RESULT 3
US-07-607-538C-4
Sequence 4, Application US/07607538C
Patent No. 5455031
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALITON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible

TITLE OF INVENTION: POLYPEPTIDE WITH 46
TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
TITLE OF INVENTION: USE THEREOF
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: V. Amzel & Assoc.
STREET: 2055 No. 5455031th Broadway
CITY: Walnut Creek
STATE: California
COUNTRY: USA
ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CRCC-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-4

Query Match 1.7%; Score 7; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 351 PRIVARY 357
|||||
Db 28 PRIVARY 34

RESULT 4
US-08-162-402B-4
Sequence 4, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALITON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible

SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, VIVIANA
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-4

Query Match 1.7%; Score 7; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 PRIVATE 357
DB 28 PRIVATE 34

RESULT 5
US-08-492-027A-4
Sequence 4, Application US/08492027A
Patent No. 591233
GENERAL INFORMATION:
APPLICANT: Suzuki, Sholchi
APPLICANT: Burnell, James N
TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/492,027A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0760-206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-492-027A-4

Query Match 1.7%; Score 7; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ALKVELI 376
DB 162 ALKVELI 168

RESULT 6
US-09-304-799-1
Sequence 1, Application US/09304799
Patent No. 6391649
GENERAL INFORMATION:
APPLICANT: The Rockefeller University
TITLE OF INVENTION: METHOD FOR THE COMPARATIVE ANALYSIS OF PROTEINS AND OTHER BIOL
FILE REFERENCE: 1119-0002
CURRENT APPLICATION NUMBER: US/09/304,799
CURRENT FILING DATE: 1999-05-04
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Version 3.0
SEQ ID NO 1
LENGTH: 108
TYPE: PRT
ORGANISM: Homo sapiens
US-09-304-799-1

Query Match 1.4%; Score 6; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 ISRYEG 196
DB 47 ISRYEG 52

RESULT 7
US-08-111-939-21
Sequence 21, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
TITLE OF INVENTION: Protein and Process for its Production
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA: JP 230029/92
FILING DATE: 28-AUG-1992

ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-21

Query Match 1.4%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 KVFQGN 338
|||||
DB 64 KVFQGN 69

RESULT 8
US-08-111-939-25
Sequence 25, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
NUMBER OF SEQUENCES: 27
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 230029/92
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-25

Query Match 1.4%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 KVFQGN 338
|||||
DB 64 KVFQGN 69

RESULT 9
US-08-111-939-23
Sequence 23, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
NUMBER OF SEQUENCES: 27
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 230029/92
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 111 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-23

Query Match 1.4%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 279 WLEIDL 284
|||||
DB 12 WLEIDL 17

RESULT 10
US-09-134-001C-4964
Sequence 4964, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4964
LENGTH: 115
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4964

Query Match 1.4%; Score 6; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 DYLFT 68
Db 73 DYLFT 78

RESULT 11
US-08-626-685A-10
Sequence 10, Application US/0862685A
Patent No. 5972624
GENERAL INFORMATION:
APPLICANT: Smith, Kelli E.
APPLICANT: Gerald, Christophe P.G.
APPLICANT: Weinschenk, Richard L.
APPLICANT: Linemeyer, David
APPLICANT: Branche, Theresa
APPLICANT: Forray, Carlos
TITLE OF INVENTION: DNA ENCODING A MAMMALIAN
TITLE OF INVENTION: HYPOTHALAMIC GALANIN RECEPTOR
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/626,685A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 50233-B/JPW/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-626-685A-10

Query Match 1.4%; Score 6; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SLCKA 168
Db 95 SLCKA 100

RESULT 12
US-08-469-260A-64
Sequence 64, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAM J. PILOT-MARTIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIK
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: POREBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-260A-64

Query Match 1.4%; Score 6; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 ASSSQ 240
Db 22 ASSSQ 27

RESULT 13
US-08-054-480-2

Sequence 2, Application US/08054480
Patent No. 5525504
GENERAL INFORMATION:
APPLICANT: Goebel, Werner
APPLICANT: Libby, Stephen
APPLICANT: Heffon, Fred
TITLE OF INVENTION: CYTOLYSIN GENE AND GENE PRODUCT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO, & BRANIGAN, P.C.
STREET: 2200 CLARENDON BOULEVARD, SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/054,480
FILING DATE: 04-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heaney, Brian P.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: MERCK 1496
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703 243 6333
TELEFAX: 703 243 6410
TELEX: 64191
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: SALMONELLA
US-08-054-480-2

Query Match 1.48; Score 6; DB 1; Length 146;
Best Local Similarity .100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 QTCASD 63
| | | | |
DB 81 QTCASD 86

RESULT 14
US-09-149-476-503
Sequence 503, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
EARLIER FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315

```

? EARLIER FILING DATE: 1997-04-11
? EARLIER APPLICATION NUMBER: 60/048,974
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/056,886
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,877
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,889
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,893
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,630
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,878
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,662
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,872
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,882
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,637
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,903
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,888
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,879
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,880
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,894
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,911
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,636
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,874
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,910
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,864
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,631
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,845
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,892
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/057,761
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/047,595
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,599
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,588
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,585
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,586
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,590
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,594
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,589
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,593
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/047,614
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/043,578
? EARLIER FILING DATE: 1997-04-11

```

```

? EARLIER APPLICATION NUMBER: 60/043,576
? EARLIER FILING DATE: 1997-04-11
? EARLIER APPLICATION NUMBER: 60/047,501
? EARLIER FILING DATE: 1997-05-23
? EARLIER APPLICATION NUMBER: 60/043,670
? EARLIER FILING DATE: 1997-04-11
? EARLIER APPLICATION NUMBER: 60/056,632
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,664
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,876
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,881
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,909
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,875
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,862
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,887
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/056,908
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/048,964
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/057,650
? EARLIER FILING DATE: 1997-09-05
? EARLIER APPLICATION NUMBER: 60/056,884
? EARLIER FILING DATE: 1997-08-22
? EARLIER APPLICATION NUMBER: 60/057,669
? EARLIER FILING DATE: 1997-09-05
? EARLIER APPLICATION NUMBER: 60/049,610
? EARLIER FILING DATE: 1997-06-13
? EARLIER APPLICATION NUMBER: 60/061,060
? EARLIER FILING DATE: 1997-10-02

```

```

Query Match 1.4%; Score 6; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 245 SGOVH 250
DB 62 SGOVH 67

```

```

RESULT 15
US-08-563-148E-6
; Sequence 6, Application US/08563148E
; Patent No. 5929224
; GENERAL INFORMATION:
; APPLICANT: Kazuo SUZUKI et al.
; TITLE OF INVENTION: NOVEL ACTIVATING FACTOR OF LEUCOCYTES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: 663-2, Shitagi
; STREET: Misaki-machi
; CITY: Isumi-gun
; STATE: Chiba-ken
; COUNTRY: JAPAN
; ZIP: 299-45
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v.5
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/563,148E
; FILING DATE: No. 5929224ember 27, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 293233/1994
; FILING DATE: 28-NOV-1994
; ATTORNEY/AGENT INFORMATION:

```

NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: intermediate fragment
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: T-cell derived leukemia cells
PUBLICATION INFORMATION:
AUTHORS: Kazuo SUZUKI et al.
TITLE: NOVEL ACTIVATING FACTOR OF LEUKOCYTES
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 to 151
US-08-563-148E-6

Query Match 1.48; Score 6; DB 2; Length 151;
Best Local Similarity 100.08; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 ISGRGF 109
|||
Db 93 ISGRGF 98

Search completed: May 15, 2003, 13:24:50
Job time : 26.9356 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:17:05 ; Search time 12.4678 Seconds

(Without alignments)
1390.548 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EEKIDGCGHLVTYODSGTWT.....STKDEDEITTRIPSEETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 89080

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	439	1 DHPC_ECOLI	P09836 escherichia
2	8	1.9	195	1 ZN90_HUMAN	003938 homo sapien
3	8	1.9	353	1 Z234_HUMAN	075437 homo sapien
4	8	1.9	418	1 Z117_HUMAN	003924 homo sapien
5	8	1.9	535	1 Z257_HUMAN	092261 homo sapien
6	8	1.9	574	1 YE73_HUMAN	092255 homo sapien
7	8	1.9	595	1 ZN85_HUMAN	003923 homo sapien
8	8	1.9	624	1 YEX7_YEAST	P40090 saccharomyc
9	7	1.7	248	1 Z2SA_BACSU	034853 bacillus su
10	7	1.7	265	1 LECN_PEA	P16270 pisum sativ
11	7	1.7	280	1 XLRL_FUGRU	096655 fugu rubrip
12	7	1.7	285	1 LEC_SOYBN	P05046 glycine max
13	7	1.7	293	1 NPL_HAETN	P45539 haemophilus
14	7	1.7	301	1 OM40_CAEEL	Q18090 caenorhabdi
15	7	1.7	322	1 Y073_LISMO	P58388 listeria in
16	7	1.7	322	1 Y016_LISMO	Q928C0 listeria in
17	7	1.7	339	1 SEHL_SCHPO	Q10099 schizosacch
18	7	1.7	356	1 CYCR_RHOVI	P07173 rhodospendo
19	7	1.7	359	1 MCAL_CRIGR	Q54873 cricetus
20	7	1.7	385	1 REGT_MYCTU	Q59386 mycobacteri
21	7	1.7	444	1 GEL2_ARATH	P49063 arabidopsis
22	7	1.7	445	1 ASSY_PASMU	P57877 pasteurella
23	7	1.7	446	1 ASSY_ECO57	Q8XMO escherichia
24	7	1.7	446	1 ASSY_ECOLI	P22767 escherichia
25	7	1.7	446	1 PRES_PSEAE	Q04804 pseudomonas
26	7	1.7	484	1 ER60_SCHMA	P38658 schistosoma
27	7	1.7	514	1 G6P1_MOUSE	Q00612 mus musculu
28	7	1.7	514	1 G6PD_MACRO	Q29492 macropus ro
29	7	1.7	514	1 G6PD_RAT	P05370 rattus norv
30	7	1.7	542	1 YM93_YEAST	P40308 saccharomyc
31	7	1.7	680	1 SYG_BOMMO	Q04451 bombyx mori
32	6	1.4	102	1 RPOL_AERPE	Q92621 aeropyrum p
33	6	1.4	109	1 RLAI_SCHPO	P17476 schizosacch

34	6	1.4	109	1 RLAI_SCHPO	Q9uu78 schizosacch
35	6	1.4	110	1 KDFG_ERWCH	Q05527 erwina chr
36	6	1.4	110	1 RLAI_SCHPO	P17477 schizosacch
37	6	1.4	114	1 MSMB_MACMU	P25142 macaca mula
38	6	1.4	114	1 YG4U_YEAST	P53308 saccharomyc
39	6	1.4	125	1 HPAA_HELIAC	Q47947 helicobacte
40	6	1.4	129	1 CY52_SCHPO	Q9uam6 schizosacch
41	6	1.4	130	1 CCRN_RANCA	P80344 rana catesb
42	6	1.4	130	1 RS8_MERTG	Q977V0 methanococc
43	6	1.4	130	1 RS8_MERTJ	P54041 methanococc
44	6	1.4	130	1 RS8_MERTL	Q977U8 methanococc
45	6	1.4	133	1 PEMK_ECOLI	P13976 escherichia

ALIGNMENTS

RESULT 1
UHPC_ECOLI STANDARD; PRT; 439 AA.
AC P09836: P76728:
DF 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Regulatory protein uhpC.
OS UHPC OR B3667.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87279903; PubMed=3301805;
RA Friedrich M.J., Kadner R.J.;
RT "Nucleotide sequence of the uhp region of Escherichia coli.";
RL J. Bacteriol. 169:3556-3563(1987).
RN [2]
RP REVISIONS.
RA Kadner R.J.;
RT Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92234930; PubMed=1569007;
RA Island M.D., Wei B.-Y., Kadner R.J.;
RT "Structure and function of the uhp genes for the sugar phosphate
transport system in Escherichia coli and Salmonella typhimurium.";
RL J. Bacteriol. 174:2754-2762(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=93315143; PubMed=766882;
RT Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [5]
RP POSSIBLE ROLE.
RA MEDLINE=93352407; PubMed=8349544;
RT Island M.D., Kadner R.J.;
RT "Interplay between the membrane-associated uhpB and uhpC regulatory
proteins.";
RL J. Bacteriol. 175:5028-5034(1993).
RT -1- FUNCTION: UHPC IS REQUIRED FOR UHPRT EXPRESSION, IT MAY ACT JOINTLY
WITH THE UHPB SIGNALING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@lsb-sib.ch.

CC EMBL; M17102; AAA24722.1; ALT_INIT.
DR EMBL; M89479; AAA24726.1; -
DR EMBL; L10328; AAA62019.1; ALT_INIT.
DR EMBL; AE000444; AAC76690.1; ALT_INIT.
DR PIR; G41853; RGEUC.
DR Ecocore; EG11053; unpc.
DR InterPro; IPR000849; GLPT_transporter.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr.1.
DR TIGRfams; TIGR00881; 2A0104; 1.
DR PROSITE; PS00942; GLPT.1.
KW Transport; Sugar transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 290 310 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
SQ SEQUENCE 439 AA; 48256 MW; D8097864B7877700 CRC64;

Query Match
Best Local Similarity 2.2%; Score 9; DB 1; Length 439;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
Db 56 ILANGVLSR 64
|||||
ID ZN90_HUMAN STANDARD; PRT; 195 AA.
AC Q03938;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 90 (Zinc finger protein HTP9) (Fragment).
GN ZNF90.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91219421; PubMed=2023309;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martiel J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; M61870; AAA36028.1; -

PIR; E39384; E39384.
DR Genew; HGNC:13165; ZNF90.
DR MIM; 603973; -
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; ZF-C2H2; 1.
DR PIR; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 4 75 KRAB.
FT ZN_FING 145 167 C2H2-TYPE (DEGENERATE).
FT ZN_FING 173 195 C2H2-TYPE.
FT NON_TER 195 195
SQ SEQUENCE 195 AA; 22562 MW; 0E808A1A62342608 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 1; Length 195;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
Db 50 PDLITCLE 57
|||||
ID Z254_HUMAN STANDARD; PRT; 353 AA.
AC O75437;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 254 (Bone marrow zinc finger 5) (BMZF-5)
DE (Hematopoietic cell derived zinc finger protein 1) (HD-ZNF1).
GN ZNF254 OR BMZF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RT "Identification of genes expressed in human CD34(+) hematopoietic
RT stem/progenitor cells by expressed sequence tags and efficient full-
RT length cDNA cloning";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20034457; PubMed=10585455;
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
RT hematopoietic cells and identification of a novel transregulatory
RT domain KRAB";
RL J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF054180; AAC39913.1; -

DR Genew: HGNC:13047; ZNF254.

DR MIM: 604768; -

DR InterPro: IPR001909; KRAB.

DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF00096; zf-C2H2; 4.

DR Pfam: PF01352; KRAB; 1.

DR PRINTS: PRO00048; ZINC_FINGER.

DR PRODOM: PD000003; Znf_C2H2; 2.

DR SMART: SM00349; KRAB; 1.

DR SMART: SM00355; Znf_C2H2; 3.

DR PROSITE: PS50805; KRAB; 1.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.

DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 4.

DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 13 84 KRAB.

FT ZN_FING 210 316 ZINC_FINGERS.

FT ZN_FING 210 232 C2H2-TYPE (DEGENERATE).

FT ZN_FING 238 260 C2H2-TYPE (DEGENERATE).

FT ZN_FING 266 288 C2H2-TYPE.

FT ZN_FING 294 316 C2H2-TYPE.

FT ZN_FING 316 316 C2H2-TYPE.

SO SEQUENCE 353 AA; 41698 MW; D923CB82D2C5B56A CRC64;

Query Match 1.9%; Score 8; DB 1; Length 353;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

Db 59 PDLITCLE 66

RESULT 4

ID 2117_HUMAN STANDARD; PRT; 418 AA.

AC 003924;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Zinc finger protein 117 (Zinc finger protein HPP9) (Fragment).

GN ZNF117.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE OF 4-418 FROM N.A.

RC TISSUE-Placenta;

RX MEDLINE=89377476; PubMed=2505992;

RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,

RA Bellefroid A., Martial J.A.,

RT "The human genome contains hundreds of genes coding for finger

RT proteins of the Kruppel type.";

RL DNA 8:377-387(1989).

RN (2)

RP SEQUENCE OF 1-166 FROM N.A.

RX MEDLINE=9119421; PubMed=2023909;

RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,

RA Martial J.A.,

RT "The evolutionarily conserved Kruppel-associated box domain defines a

RT subfamily of eukaryotic multifingered proteins.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).

CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.

CC -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M27879; -; NOT ANNOTATED_CDS.

DR EMBL: M61867; AA58666.1; -

DR PIR: B39384; B39384.

DR HSP: P03001; 1TF6.

DR Genew: HGNC:12897; ZNF117.

DR MIM: 194624; -

DR InterPro: IPR001909; KRAB.

DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF00096; zf-C2H2; 10.

DR PRODOM: PD000003; Znf_C2H2; 5.

DR SMART: SM00349; KRAB; 1.

DR SMART: SM00355; Znf_C2H2; 10.

DR PROSITE: PS50805; KRAB; 1.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.

DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.

DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 1 46 KRAB.

FT ZN_FING 116 417 ZINC_FINGERS.

FT ZN_FING 116 138 C2H2-TYPE (DEGENERATE).

FT ZN_FING 144 166 C2H2-TYPE.

FT ZN_FING 172 194 C2H2-TYPE (DEGENERATE).

FT ZN_FING 200 222 C2H2-TYPE.

FT ZN_FING 228 250 C2H2-TYPE.

FT ZN_FING 256 278 C2H2-TYPE.

FT ZN_FING 284 306 C2H2-TYPE.

FT ZN_FING 312 334 C2H2-TYPE.

FT ZN_FING 340 362 C2H2-TYPE.

FT ZN_FING 368 390 C2H2-TYPE (DEGENERATE).

FT ZN_FING 396 417 C2H2-TYPE.

FT CONFLICT 75 75 R -> G (IN REF. 2).

FT CONFLICT 78 79 G -> RH (IN REF. 2).

SO SEQUENCE 418 AA; 48941 MW; CDB554CB69B0B69 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 418;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

Db 21 PDLITCLE 28

RESULT 5

ID 2257_HUMAN STANDARD; PRT; 535 AA.

AC Q9Y201;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).

GN ZNF257 OR BMZF4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE-Bone marrow;

RX MEDLINE=20054457; PubMed=10585455;

RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,

RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.,

RT "Molecular cloning of six novel Kruppel-like zinc finger genes from

RP hematopoietic cells and identification of a novel transregulatory
domain KRAB.";
RL J. Biol. Chem. 274:35741-35748 (1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF070651; AAD20957.1; -
DR HSSP; P08047; 1SR2.
DR Genew; HGNC:13498; ZNF257.
DR MIM; 606957; -
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 10.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 9.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 10.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
RW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 4 75 KRAB.
FT ZN_FING 173 509 ZINC_FINGERS.
FT ZN_FING 173 195 C2H2-TYPE.
FT ZN_FING 201 223 C2H2-TYPE.
FT ZN_FING 229 251 C2H2-TYPE.
FT ZN_FING 257 282 C2H2-TYPE (DEGENERATE).
FT ZN_FING 288 310 C2H2-TYPE.
FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
FT ZN_FING 459 481 C2H2-TYPE.
FT ZN_FING 487 509 C2H2-TYPE.
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;
Query Match 1.9%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY .119 PDLITCLE 126
DB 50 PDLITCLE 57
RESULT 6
YE73_HUMAN STANDARD: PRT; 574 AA.
ID YE73_HUMAN
AC Q9P235;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA1473 (Fragment).
CN KIAA1473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
GN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB040906; BAA95997.1; -
DR HSSP; P07248; 1ARD.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 13.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 2.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.
RW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT NON_TER 1 1
FT DOMAIN 15 86 KRAB.
FT ZN_FING 184 542 ZINC_FINGERS.
FT ZN_FING 184 206 C2H2-TYPE.
FT ZN_FING 212 234 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 290 C2H2-TYPE.
FT ZN_FING 296 318 C2H2-TYPE.
FT ZN_FING 324 346 C2H2-TYPE.
FT ZN_FING 352 374 C2H2-TYPE.
FT ZN_FING 380 402 C2H2-TYPE.
FT ZN_FING 408 430 C2H2-TYPE.
FT ZN_FING 436 458 C2H2-TYPE.
FT ZN_FING 464 486 C2H2-TYPE.
FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
FT ZN_FING 520 542 C2H2-TYPE.
SQ SEQUENCE 574 AA; 65951 MW; 482E5F1176BCA5B0 CRC64;
Query Match 1.9%; Score 8; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY .119 PDLITCLE 126
DB 61 PDLITCLE 68
RESULT 7
ZN85_HUMAN STANDARD: PRT; 595 AA.
ID ZN85_HUMAN
AC Q03923;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 85 (zinc finger protein HRP4) (HNF1).
GN ZNF85.

OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_Taxid=9606;
 RX [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC TISSUE-PLACENTA;
 RX MEDLINE=99053537; PubMed=9839802;
 RA Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demotie M.A.,
 RA Marlier J.C., Penderville H., Alam Y., Devos N., Lecocq P.J., Ogawa T.,
 RA Muller M., Marlier J.A.;
 RT "functional analysis of ZNF85 KRAB zinc finger protein, a member of
 RT the highly homologous ZNF91 family.";
 RL DNA Cell Biol. 17:931-943(1998).
 RN [2]
 RP SEQUENCE OF 1-196 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Marlier J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN TESTICULAR
 CC TISSUES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U35376; AAA79179.1; -
 DR EMBL: M61866; AAA52689.1; -
 DR EMBL: M61868; AAA58671.1; -
 DR PIR: A39384; A39384.
 DR PIR: C39384; C39384.
 DR HSSP: P08048; ZNF.
 DR TRANSFAC: T04990; -
 DR Gene: HGNC:13160; ZNF85.
 DR MIM: 603899; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; ZNF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 15.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINCFINGER.
 DR ProDom: PD000003; Znf.C2H2; 13.
 DR SMART: SM00349; KRAB_1;
 DR SMART: SM00355; ZNF_C2H2; 15.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 15.
 DR Transcription regulation: Zinc-finger; DNA-binding; Metal-binding;
 KM Nuclear protein; Repeat; Repressor.
 FT DOMAIN 4 75 KRAB.
 FT ZINC FINGERS.
 FT ZN_FING 146 588 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 146 168 C2H2-TYPE.
 FT ZN_FING 174 196 C2H2-TYPE.
 FT ZN_FING 202 224 C2H2-TYPE.
 FT ZN_FING 230 252 C2H2-TYPE.
 FT ZN_FING 258 280 C2H2-TYPE.
 FT ZN_FING 286 308 C2H2-TYPE.
 FT ZN_FING 314 336 C2H2-TYPE.
 FT ZN_FING 342 364 C2H2-TYPE.
 FT ZN_FING 370 392 C2H2-TYPE.
 FT ZN_FING 398 420 C2H2-TYPE.

FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 454 476 C2H2-TYPE.
 FT ZN_FING 482 504 C2H2-TYPE.
 FT ZN_FING 510 532 C2H2-TYPE.
 FT ZN_FING 538 560 C2H2-TYPE.
 FT ZN_FING 566 588 C2H2-TYPE.
 FT CONFLICT 84 84 R -> Q (IN REF. 2; AAA52689).
 FT CONFLICT 115 115 R -> I (IN REF. 2; AAA58671).
 FT CONFLICT 177 177 T -> R (IN REF. 2; AAA58671).
 FT CONFLICT 184 184 G -> R (IN REF. 2; AAA58671).
 SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D5D3B CRC64;
 QY 119 PDITCILE 126 1.9%; Score 8; DB 1; Length 595;
 DB 50 PDITCILE 57 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 8
 YEST_YEAST STANDARD; PRT; 624 AA.
 ID YEST_YEAST
 AC P40090;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 72.1 kDa protein in FTR1-SPT15 intergenic region.
 GN YER147C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Ayiles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Kosdale D., Nakahara K., Namath A., Norgren R., Oetner P., Oh C.,
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U18917; AAB6474.1; -
 DR SGD: S0000949; YER147C.
 DR InterPro: IPR000873; AMP-bind.
 DR PROSITE: PS00455; AMP_BINDING; 1.
 KM Hypothetical protein; Ligase.
 SQ SEQUENCE 624 AA; 72140 MW; 1269D931C656608 CRC64;
 QY 368 RIALKVEL 375 1.9%; Score 8; DB 1; Length 624;
 DB 285 RIALKVEL 292 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 9

DR PROSITE; PS00307; LECTIN_LEGUME_BETA; FALSE_NEG.
 DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; FALSE_NEG.
 KW LECTIN; Signal; Multigene family; Manganese; Calcium; Glycoprotein.
 FT CHAIN 1 21 OR 23 (POTENTIAL).
 FT SIGNAL 22 25
 FT CARBOHYD 59 59 NONSEED LECTIN.
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 265 AA; 28530 MW; BC3CC9BD7DCAB82 CRC64; (POTENTIAL).
 Query Match 1.7%; Score 7; DB 1; Length 265;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 197 ILANGVL 203
 Db 49 ILANGVL 55
 RESULT 11
 ID XLR1_FUGRU STANDARD; PRT; 280 AA.
 AC Q9W6R5;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE X-linked juvenile retinosis protein precursor.
 GN XLR1.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 RN NCBI_TaxID=31033;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99263230; PubMed=10330123;
 RA Brunner B., Todd T., Lenzner S., Stout K., Schulz U., Ropers H.-H.,
 RA Kalscheuer V.M.;
 RT "Genomic structure and comparative analysis of nine Fugu genes:
 RT conservation of synteny with human chromosome Xp22.2-p22.1.";
 RL Genome Res. 9:437-448(1999).
 CC -1- FUNCTION: MAY BE ACTIVE IN CELL ADHESION PROCESSES DURING RETINAL
 CC DEVELOPMENT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AF146687; AAD28797.1; -;
 DR HSSP; P12259; 1CZT.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR SMART; SM00231; FA58C_1; 1.
 DR PROSITE; PS01285; FA58C_1; 1.
 DR PROSITE; PS01286; FA58C_2; FALSE_NEG.
 KW Cell adhesion; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 26 POTENTIAL.
 FT DOMAIN 119 275 F5/8 TYPE C.
 FT SIGNAL 119 275 BY SIMILARITY.
 SQ SEQUENCE 280 AA; 32032 MW; A0DF9A3222ED0167 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 280;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 PPIVARY 357
 Db 248 PPIVARY 254
 RESULT 12
 ID LEC_SOYBN STANDARD; PRT; 285 AA.
 AC P05046;
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-JUL-1999 (Rel. 38, Last annotation update)
 DE Lectin precursor (Agglutinin) (SBA).
 GN LEL.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 RN NCBI_TaxID=3847;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84026469; PubMed=6313203;
 RA Vodka L.O., Rhodes P.R., Goldberg R.B.;
 RT "A lectin gene insertion has the structural features of a
 RT transposable element.";
 RL Cell 34:1023-1031(1983).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95226380; PubMed=7711015;
 RA Dessen A., Gupta D., Sabesan S., Brewer C.F., Sacchetti J.C.;
 RT "X-ray crystal structure of the soybean agglutinin cross-linked with
 RT a biannenary analog of the blood group I carbohydrate antigen.";
 RL Biochemistry 34:4933-4942(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.52 ANGSTROMS).
 RX MEDLINE=98060759; PubMed=9398234;
 RA Olsen L.R., Dessen A., Gupta D., Sabesan S., Sacchetti J.C.,
 RA Brewer C.F.;
 RT "X-ray crystallographic studies of unique cross-linked lattices
 RT between four isomeric biannenary oligosaccharides and soybean
 RT agglutinin.";
 RL Biochemistry 36:15073-15080(1997).
 CC -1- FUNCTION: BINDS GLCNAC AND GALACTOSE.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SIMILARITY: BELONGS TO THE LEGUMINOS LECTIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; K00821; AAA33983.1; -;
 DR PIR; S27365; S27365.
 DR PDB; 2SBA; 09-DEC-98.
 DR PDB; 1SBD; 22-APR-98.
 DR PDB; 1SBE; 22-APR-98.
 DR PDB; 1SBE; 22-APR-98.
 DR GlycoSuiteDB; P05046; -;
 DR InterPro; IPR000985; LECTIN_LEG.
 DR InterPro; IPR001220; LECTIN_LEG.
 DR Pfam; PF00139; LECTIN_LEG; 1.
 DR ProDom; PD000671; LECTIN_LEG; 1.
 DR ProDom; PD000711; LECTIN_LEG; 1.
 DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
 DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
 KW LECTIN; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 32
 FT CHAIN 33 285 LECTIN.
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .).

SQ SEQUENCE 285 AA; 30928 MW; B3704533C9315C52 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 285;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 NESGDY 249
 |||||
 Db 146 NESGDY 152

RESULT 13
 NPL_HAEIN STANDARD; PRT; 293 AA.
 ID NPL_HAEIN
 AC P44539;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable N-acetylneuraminic lyase subunit (EC 4.1.3.3) (N-acetylneuraminic acid aldolase) (N-acetylneuraminic pyruvate lyase) (Nalase).
 GN NANA OR H10142.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 NCBI_TaxID=727;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-Rd / KM20 / ATCC 51907;
 RX MEDLINE=9530630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Saiton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shrivley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uettermann T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., Greim C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae R.";
 RT Science 269:496-512(1995).
 CC -1- CATALYTIC ACTIVITY: N-acetylneuraminic acid -> N-acetyl-D-mannosamine + pyruvate.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DHPS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U37700; AAC21814.1; -.
 DR HSSP: P06995; INAL.
 DR TIGR: H10142; -.
 DR InterPro: IPR002220; DHDS.
 DR InterPro: IPR005264; Nana.
 DR Pfam: PF00701; DHDS.1.
 DR PRINTS: PR00146; DHPICNTHASE.
 DR PRODOM: PD001859; DHDS.1.
 DR TIGRFAMs: TIGR00683; nana.1.
 DR PROSITE: PS00665; DHDS.1; 1.
 DR PROSITE: PS00666; DHDS.2; 1.
 KM Lyase: Complete proteome.
 FT ACT_SITE 164 164 BY SIMILARITY.
 SQ SEQUENCE 293 AA; 32564 MW; ACAV550DA5D93D33 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 293;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 EGIANG 201
 |||||
 Db 243 EGIANG 249

RESULT 14
 OM40_CAEEL STANDARD; PRT; 301 AA.
 ID OM40_CAEEL
 AC Q18090;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable mitochondrial import receptor subunit TOM40 homolog (Translocase of outer membrane 40 kDa subunit homolog).
 GN C18B3.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
 OC Rhabditidae; Rhabditidae; Caenorhabditis.
 NCBI_TaxID=6239;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2.
 RA Sims M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL FOR THE IMPORT OF PROTEIN PRECURSORS INTO THE MITOCHONDRION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial outer membrane (BY similarity).
 CC -1- SIMILARITY: BELONGS TO THE TOM40 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z70034; CAA93850.1; -.
 DR Wormpep: C18B3.6; CE05298.
 KW Hypothetical protein; Transport; Protein transport; Outer membrane; Mitochondrion; Transmembrane.
 KM Mitochondrion; Transmembrane.
 SQ SEQUENCE 301 AA; 32386 MW; 4247A3593B5ECC57 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 301;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 GGIISVL 186
 |||||
 Db 181 GGIISVL 187

RESULT 15
 Y073_LISMO STANDARD; PRT; 322 AA.
 ID Y073_LISMO
 AC P58588;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein lmo2473.
 GN LMO2473.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 NCBI_TaxID=1639;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangeul L., Buchrieser C., Rusnok C., Amend A., Bagueri F., Berche P., Bloeker H., Brandt P., Chakraborty T., Chardot A., Chetoui F., Couve E., de Darvar A., Denoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.D., Esih H., Portillo F.G., Garrido P., Gautier L.,
 RA Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.M.,
 RA Kaerst U., Krett J., Kuhn M., Kunst F., Kurapkat G., Madueno E.,
 RA Maitournam A., Vicente J.M., Ng E., Nedjari H., Nordstiek G.,
 RA Novella S., de Pablos B., Perez-Diaz J.C., Purcell R., Rammel B.,
 RA Rose M., Schluter T., Simoes N., Tierrez A., Vazquez-Boland J.A.,
 RA Voss H., Wehlund J., Cossart P.;
 RT *Comparative genomics of *Listeria* species.*;
 RL Science 294:849-852(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0052 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL591983; CAD00551.1; -.
 DR Listlist; LMO02473; -.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 322 AA; 34973 MW; 50180FCBADF63A31 CRC64;
 QY 82 TVPKELL 88
 DB 258 TVPKELL 264
 Query Match 1.7%; Score 7; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: May 15, 2003, 13:23:05
 Job time : 16.4678 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:16:36 ; Search time 22.3469 Seconds
(without alignments)
1305.856 Million cell updates/sec

Title: US-10-003-132-4-COPY_26_244

Perfect score: 219
Sequence: 1 EELDGCGHIVTSQDSGTWT.....NGVLSRHGSLSEKRELFETTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 306882

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: listing first 45 summaries

Database :

1: A.Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	67.6	503	21	AA19126
2	28	12.8	487	22	AAU00628
3	28	12.8	539	22	AAU00630
4	28	12.8	586	22	AAU00629
5	20	9.1	398	23	AAE22715
6	20	9.1	398	23	AAU79459
7	8	3.7	101	22	AAU19467
8	8	3.7	101	23	ABP51369
9	8	3.7	101	23	ABP51461
10	8	3.7	109	22	AA638946

11	8	3.7	117	22	AA92967
12	8	3.7	117	23	ABP51401
13	8	3.7	128	22	AA79595
14	8	3.7	193	22	AA40732
15	8	3.7	273	22	ABG6004
16	8	3.7	353	20	AA139779
17	8	3.7	364	22	AA94710
18	8	3.7	493	22	ABG16286
19	8	3.7	694	21	AA99661
20	7	3.2	115	22	ABG27684
21	7	3.2	126	22	AA90976
22	7	3.2	204	23	AA48377
23	7	3.2	206	21	AA96283
24	7	3.2	217	23	ABP26698
25	7	3.2	220	22	ABG21553
26	7	3.2	263	15	AA63754
27	7	3.2	275	22	AA86528
28	7	3.2	297	22	ABG12597
29	7	3.2	312	22	AAU14605
30	7	3.2	350	21	AA812145
31	7	3.2	350	22	AAW93217
32	7	3.2	350	22	AA93381
33	7	3.2	350	22	AA94684
34	7	3.2	359	20	AA81359
35	7	3.2	385	22	AA681097
36	7	3.2	414	22	AAE10702
37	7	3.2	430	23	ABP28272
38	7	3.2	464	22	AAW93631
39	7	3.2	474	22	AB62780
40	7	3.2	474	22	AB66283
41	7	3.2	523	18	AAW38187
42	7	3.2	524	21	AA828588
43	7	3.2	610	22	AAU34821
44	7	3.2	639	22	ABG17654
45	7	3.2	641	22	AB58034

ALIGNMENTS

RESULT 1	AA19126
ID	AA19126 standard; Protein; 503 AA.
XX	AA19126;
AC	AA19126;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Polypeptide isolated, from lymph node stromal cells of fsn -/- mice.
XX	
KW	Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW	immune system disorder; cancer; viral infection;
KW	blood vessel growth; tumour necrosis factor disorder; arthritis;
KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW	cardiac failure.
XX	
OS	Mus sp.
XX	
PN	W0200058463-A1.
XX	
PD	05-OCT-2000.
XX	
PF	18-FEB-2000; 2000WO-N200015.
XX	
PR	25-MAR-1999; 99US-0276268.
XX	
PR	26-AUG-1999; 99US-0383586.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	
PI	Strachan L, Sleeman M, Abernethy N, Orrust R, Kumble KD;
XX	Murison JG;
DR	WPI; 2000-664924/64.

PT or diseases -
XX
PS Claim 5; Page 29-30; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
CC
SQ Sequence 539 AA;
Query Match 12.8%; Score 28; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 4e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
DB 134 GSGHISGRGFLTYASSDHPDLITCLER 161
|||||
RESULT 4
AA000629
ID AA000629 standard; Protein; 586 AA.
XX
AC AA000629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000MO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00614.
XX
XX Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
XX Claim 2; Page 27-28; 33pp; English.
PS The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing

CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
CC
SQ Sequence 586 AA;
Query Match 12.8%; Score 28; DB 22; Length 586;
Best Local Similarity 100.0%; Pred. No. 4.3e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
DB 181 GSGHISGRGFLTYASSDHPDLITCLER 208
|||||
RESULT 5
AAE22715
ID AAE22715 standard; Protein; 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW ankyrotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW noctropic; neuroprotective; vulnerary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
PN WO200222815-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001MO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX WPI; 2002-393966/42.
DR N-PSDB; AAD35992.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
XX
XX Claim 3; Page 123-125; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, plaitic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders. Infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye diseases. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy1
CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

SO Sequence 398 AA;

Query Match 9.1%; Score 20; DB 23; Length 398;

Best Local Similarity 100.0%; Pred. No. 2.6e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SSGSHSGRGFLTYASSDHP 119

100 SSGSHSGRGFLTYASSDHP 119

RESULT 6
ID AAU79459 standard; Protein: 398 AA.
AC AAU79459;

15-JUL-2002 (first entry)

Human Neuropilin-Hy1.

Human; neuropilin-Hy1; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.

Homo sapiens.

WO200222780-A2.

21-MAR-2002.

11-SEP-2001; 2001WO-US28590.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-0659671.

(TANG/) TANG T Y.

PI Tang TY;
XX WPI: 2002-351881/38.
DR N-PSDB: ABR49565.

PT New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation
PS Claim 3; Page 118-120; 144pp; English.

CC The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy1.

SO Sequence 398 AA;

Query Match 9.1%; Score 20; DB 23; Length 398;

Best Local Similarity 100.0%; Pred. No. 2.6e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SSGSHSGRGFLTYASSDHP 119

100 SSGSHSGRGFLTYASSDHP 119

RESULT 7
ID AAU19467 standard; Protein: 101 AA.
AC AAU19467;

04-DEC-2001 (first entry)

Human diagnostic and therapeutic polypeptide (DITRP) #53.

Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder.

Homo sapiens.

WO200162927-A2.

XX 30-AUG-2001.
 PD 21-FEB-2001; 2001MO-US06059.
 XX
 XX 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 XX Chen A, D'Sa SA, Amshy S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiter TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
 XX
 DR WPI; 2001-502867/55.
 DR N-PSDB; AAS31038.
 XX
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics.
 PT
 PT
 XX Claim 27; Page 430; 522pp; English.
 PS
 XX The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DIRNP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DIRNP expression. For example, (I) and
 CC (II) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DIRNPs, by expressing inactive proteins
 CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
 CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DIRNPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative

CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DIRNPs and in assays to identify modulators of DIRNP
 CC expression and activity. The anti-DIRNP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DIRNP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DIRNPs in samples (e.g. by enzyme linked immunosorbent
 CC assay (ELISA)). AA019415-AA019625 represent human diagnostic and
 CC therapeutic (DIRNP) polypeptides of the invention.
 XX
 XX Sequence 101 AA;
 QY 119 PDLITCLE 126
 Db 59 PDLITCLE 66
 Query Match 3.7%; Score 8; DB 22; Length 101;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 RESULT 8
 ID ABB51369 standard; Protein; 101 AA.
 AC ABB51369;
 DE 03-SEP-2002 (first entry)
 DE Human MDDT SEQ ID NO 391.
 XX
 XX Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW Proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatocytic; antiinflammatory; antiproliferic; cytostatic; anti-HIV;
 KW antiallergic; antianemic; antisthmatic; antithrombotic; antitumor;
 KW neuroprotective; antineuritic; antineuritic; antineuritic.
 XX
 OS Homo sapiens.
 XX
 XX WO200240715-A2.
 PD 23-MAY-2002.
 XX
 PF 06-SEP-2001; 2001MO-US27628.
 XX
 XX 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230559P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230656P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
 PI Komiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gersting EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 XX
 DR WPI; 2002-527544/56.

DR N-PSDB; ABQ72586.
 XX
 PR Novel human disease detection and treatment polypeptide, useful in
 PR diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PR arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PR e.g. AIDS
 PS Claim 14; Page 541-542; 618pp; English.
 XX
 XX The invention relates to an isolated human disease detection and
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDP in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDP in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition.
 CC associated with decreased or increased expression of functional MDP.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDP, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knock-in humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.
 XX
 SQ Sequence 101 AA;
 Query Match 3.7%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 119 PDLITCLE 126
 |||||||
 Db 59 PDLITCLE 66
 RESULT 9
 ABP51461
 ID ABP51461 standard; Protein; 101 AA.
 AC
 AC ABP51461;
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDP ID NO 483.
 XX
 XX Human; MDP; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytoskeletal; anti-HIV;
 KW antiallergic; antianaemic; antiasumatic; antiatherosclerotic; anti-gout;
 KW neuroprotective; antirheumatic; antiarthritic.
 KW
 XX Homo sapiens.
 OS
 XX WO2002040715-A2.

XX
 PD 23-MAY-2002.
 XX
 PF 06-SEP-2001; 2001WO-US27628.
 XX
 PR 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230655P.
 PR 06-SEP-2000; 2000US-230888P.
 PR 06-SEP-2000; 2000US-230899P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Gletzen D, Liu TF, Yap PE, Dahl CR;
 PI Momiyama NG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerlitz EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
 XX
 DR WPI; 2002-527544/56.
 DR N-PSDB; ABQ72678.
 XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS
 PS Claim 14; Page 603; 618pp; English.
 XX
 XX The invention relates to an isolated human disease detection and
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDP in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDP in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition.
 CC associated with decreased or increased expression of functional MDP.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDP, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knock-in humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.
 XX

SQ Sequence 101 AA;
 Query Match 3.7%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||||
 DB 59 PDLITCLE 66
 RESULT 10
 AAM38946
 ID AAM38946 standard; Protein; 109 AA.
 AC AAM38946;
 XX
 XX 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 2091.
 DE Human: nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CMS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 XX
 XX WO200153312-A1.
 PN
 PD 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000MO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI58102.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 3; SEQ ID NO 2091; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 109 AA;
 Query Match 3.7%; Score 8; DB 22; Length 109;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||||
 DB 82 PDLITCLE 89
 RESULT 11
 AAB92967
 ID AAB92967 standard; Protein; 117 AA.
 AC AAB92967;
 XX
 XX 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:11667.
 DE Human: primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS Homo sapiens.
 XX
 XX EP1074617-A2.
 PN
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 28-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI
 PT full-length cDNAs -
 PT full-length cDNAs -
 XX
 DR WPI: 2001-318749/34.
 PS Claim 8; SEQ ID-11667; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SO Sequence 117 AA;

Query Match 3.7%; Score 8; DB 22; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||||
DB 81 PDLITCLE 88

RESULT 12

ABP51401
ID ABP51401 standard; Protein: 117 AA.

AC ABP51401;

DT 03-SEP-2002 (first entry)

DE Human MDDF SEQ ID NO 423.

XX Human; MDDF; disease detection and treatment molecule polynucleotide;
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; arteriosclerosis;
KW hepatotropic; antiinflammatory; antipruritic; cycostatic; anti-HIV;
KW antiallergic; antianemic; antistatic; antithrombotic; antileukemic; antineoplastic;
KW neuroprotective; antineuritic; antineuritic.

XX Homo sapiens.

OS WO200240715-A2.

PN 23-MAY-2002.

PD 06-SEP-2001; 2001WO-US27628.

XX 06-SEP-2000; 2000US-230505P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230655P.
PR 06-SEP-2000; 2000US-230988P.
PR 06-SEP-2000; 2000US-230989P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.

XX (INCY-) INCYTE GENOMICS INC.

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JT;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX WPI; 2002-527544/56.
DR N-PSDB; ABO72618.

PT Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

PT e.g. AIDS -
XX Claim 14; Page 562; 618pp; English.

CC The invention relates to an isolated human disease detection and
CC treatment (MDDF) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDDF in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDDF in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDDF.
CC (I) or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDDF, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases. In somatic or
CC germ-line gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.

SO Sequence 117 AA;

Query Match 3.7%; Score 8; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
|||||||
DB 90 PDLITCLE 97

RESULT 13

AAM79595
ID AAM79595 standard; Protein: 128 AA.

AC AAM79595;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3241.
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
OS WO200157190-A2.
PN 09-AUG-2001.
PD 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-046914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0726422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB; AAK52728.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 293; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAH80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 128 AA;
XX
Query Match 3.7%; Score 8; DB 22; Length 128;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 119 PDLITCLE 126
Db 101 PDLITCLE 108
XX
RESULT 14
AAM40732
ID AAM40732 standard; Protein; 193 AA.
XX
AC AAM40732;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5663.
XX
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0486725.
PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AAI59888.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5663; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAH42213) with neotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 193 AA;
XX
Query Match 3.7%; Score 8; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 119 PDLITCLE 126
Db 100 PDLITCLE 107
XX
RESULT 15
ABG60044
ID ABG60044 standard; Protein; 273 AA.
XX
AC ABG60044;
XX
DT 30-JUL-2002 (first entry)
XX
DE Human DITHP polypeptide #102.
XX
KW Human; DITHP; diagnostic and therapeutic polypeptide; bone; testis; skin;
KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
KW inflammatory disorder; viral infection; bacterial infection; seizure;
KW fungal infection; parasitic infection; developmental disorder; breast;
KW endocrine disorder; metabolic disorder; neurological disorder; cervix;
KW gastrointestinal disorder; transport disorder; gene therapy; kidney;
KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
KW thymus.
XX
OS Homo sapiens.
XX
PN WO200220754-A2.
XX
PD 14-MAR-2002.

```

XX 29-AUG-2001; 2001WO-US27127.
PF
XX
PR 05-SEP-2000; 2000US-229747P.
PR 05-SEP-2000; 2000US-229748P.
PR 05-SEP-2000; 2000US-229749P.
PR 05-SEP-2000; 2000US-229750P.
PR 05-SEP-2000; 2000US-229751P.
PR 05-SEP-2000; 2000US-230583P.
PR 05-SEP-2000; 2000US-230585P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230655P.
PR 06-SEP-2000; 2000US-230988P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Datto A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
XX
DR WPI: 2002-383054/41.
DR N-PSDB: ABK71635.
XX
XX
PT An isolated polynucleotide useful in diagnostics and therapeutics -
PS
PS Claim 29; Page 588-589; 686pp; English.
XX
XX
CC The invention relates to human diagnostic and therapeutic (dthp)
CC polynucleotides and their associated polypeptides (DTHP polypeptides).
CC The sequences of the invention are used in the treatment and diagnosis of
CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers
CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,
CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or
CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,
CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal
CC infections, parasitic infections, developmental disorders (e.g. anaemia,
CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),
CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders
CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders
CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences
CC ABG59943-ABG60220 represent human DTHP polypeptides of the invention.
XX
SQ Sequence 273 AA:

```

```

Query Match 3 7%: Score 8; DB 23; Length 273;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 119 PDLITCLE 126
DB 101 PDLITCLE 108

```

```

Search completed: May 15, 2003, 13:22:40
Job time : 23.3469 secs

```

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:21:41 ; Search time 19.2527 Seconds
(without alignments)
1093.529 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244
Perfect score: 219
Sequence: 1 EEELGDCGHIWTSQDSGTMT.....NGVLSRHGSLSEKRFLETPP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 223049

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

- 1: PIR1:**
- 2: PIR2:**
- 3: PIR3:**
- 4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.7	154	2 D39384	finger protein HTF
2	8	3.7	163	2 B39384	finger protein HTF
3	8	3.7	195	2 E39384	finger protein HTF
4	8	3.7	196	2 A39384	finger protein HTF
5	8	3.7	440	1 RGEUC	regulatory protein
6	8	3.7	440	2 D91204	regulator of uHPT
7	8	3.7	440	2 F86050	transducin homolog
8	8	3.7	519	2 A49367	transcription repr
9	8	3.7	595	2 G02075	probable ATP /GTP
10	8	3.7	135	2 DB1337	hypothetical prote
11	8	3.7	152	2 JE0031	hypothetical prote
12	8	3.7	182	2 T34009	riboflavin synthas
13	8	3.7	203	2 AG2720	beta-lactamase (EC
14	8	3.7	204	2 S51379	conserved hypochet
15	8	3.7	291	2 JP0074	phosphoribosylglyc
16	8	3.7	302	2 T50737	recf protein - Myc
17	8	3.7	314	2 C90256	keratin, type I cy
18	8	3.7	320	2 T27372	cysteine proteinas
19	8	3.7	384	1 E69685	hypothetical prote
20	8	3.7	385	2 S70984	hypothetical prote
21	8	3.7	419	2 A25438	scarcrow-like pro
22	8	3.7	494	2 S51117	hypothetical prote
23	8	3.7	507	2 T00682	hypothetical prote
24	8	3.7	573	2 T51239	hypothetical prote
25	8	3.7	578	2 B71431	probable phosphos
26	8	3.7	578	2 S51379	DNA helicase reco
27	8	3.7	610	1 BVECR0	ATP-dependent DNA
28	8	3.7	611	2 H91222	ATP-dependent DNA
29	8	3.7	611	2 F86069	ATP-dependent DNA

30	7	3.2	642	2 S59306	probable membrane
31	6	2.7	102	2 A72739	probable DNA-direc
32	6	2.7	110	2 S17709	kdgF protein - Btw
33	6	2.7	111	2 T40753	very hypothetical
34	6	2.7	117	2 T35913	hypothetical prote
35	6	2.7	122	2 A84989	50S ribosomal prot
36	6	2.7	123	2 AC2172	hypothetical prote
37	6	2.7	123	2 AC2259	hypothetical prote
38	6	2.7	126	2 S30707	hypothetical 14.5k
39	6	2.7	126	2 AE0918	probable membrane
40	6	2.7	127	2 B85638	hypothetical prote
41	6	2.7	129	2 G90761	hypothetical prote
42	6	2.7	129	2 B85625	hypothetical prote
43	6	2.7	132	2 A82847	hypothetical prote
44	6	2.7	134	2 E75356	hypothetical prote
45	6	2.7	137	1 PWYCE	H+-transporting tw

ALIGNMENTS

RESULT 1
D39384
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
C:Accession: D39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kriepel-associated box domain defines a subfam
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: D39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <BRL>
A:Cross-references: GB:M61869; NID:g184449; PID:g184450
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 8 PDLITCLE 15

RESULT 2
B39384
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 08-Dec-2000
C:Accession: B39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kriepel-associated box domain defines a subfam
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: B39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <AAA>
A:Cross-references: GB:M61867
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 18 PDLITCLE 25

RESULT 3
E39384
finger protein Htf9 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 30-Jan-1993 #text_change 08-Dec-2000
C:Accession: E39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kruppel-associated box domain defines a subfamily
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: E39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <BEL>
A:Cross-references: GB:M61869
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 50 PDLITCLE 57

RESULT 4
A39384
finger protein HPR4 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
C:Accession: A39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kruppel-associated box domain defines a subfamily
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: A39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <BEL>
A:Cross-references: GB:M61866; MUID:9454818; PIDN:AAA52689.1; PID:9184336
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 50 PDLITCLE 57

RESULT 5
RGECUC
regulatory protein uhpc - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 30-Jun-1991 #sequence_revision 14-Nov-1997 #text_change 01-Mar-2002
C:Accession: D65168; G41853; C26925; S30078
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D65168
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-440 <BLAT>
A:Cross-references: GB:AE000444; GB:U00096; MUID:92367258; PIDN:AAC76690.1; PID:92367260;
A:Experimental source: strain K-12, substrain MG1655
J. Bacteriol. 174, 2754-2762, 1992

A:Title: Structure and function of the uhpc genes for the sugar phosphate transport sy
A:Reference number: A41853; MUID:92234930; PMID:1569007
A:Accession: G41853
A:Molecule type: DNA
A:Residues: 2-440 <ISL>
A:Cross-references: GB:M69479; MUID:9148116; PIDN:AAA24726.1; PID:9148119
A:Note: This is a revision to the sequence from reference A30395
R:Frederich, M.J.; Kadner, R.J.
J. Bacteriol. 169, 3556-3563, 1987
A:Title: Nucleotide sequence of the uhpc region of Escherichia coli.
A:Reference number: A30395; MUID:87279903; PMID:3301805
A:Accession: C26925
A:Molecule type: DNA
A:Residues: 'M', '87', 'VCRAICAE', '96', 'L', '98', 'P', '100', 'L', '102', 'FTG', '139-141', 'SVNGLVFTYRA', '1
A:Note: This sequence has been revised in reference A41853
C:Comment: This is one of the proteins involved in the expression of uhpc, a gene for
MD transmembrane orientation.

Query Match 3.7%; Score 8; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 LANGVLSR 205
DB 58 LANGVLSR 65

RESULT 6
D91204
regulator of uhpc [imported] - Escherichia coli (strain O157:H7, substrain RMD 05099
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D91204
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: D91204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA38027.1; PID:913364079; GSFDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS4604
C:Superfamily: hexose phosphate transport protein uhpc

Query Match 3.7%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 LANGVLSR 205
DB 58 LANGVLSR 65

RESULT 7
F86050
regulator of uhpc [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F86050
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Ntute 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F86050
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-440 <STO>
 A:Cross-references: GB:AE005174; NID:912518499; PIDN:AAG5866.1; GSPDB:GN00145; OMGP:Z51
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: uhpC
 C:Superfamily: hexose phosphate transport protein uhpT

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 440;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 LANGVLSR 205
 |||||
 Db 58 LANGVLSR 65

RESULT 8
 A49367
 transducin homolog sazd - human
 C:Species: Homo sapiens (man)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-May-2000
 C:Accession: A49367
 R:Weinstat-Saslow, D.L.; Germino, G.G.; Somlo, S.; Reeders, S.T.
 A:Title: A transducin-like gene maps to the autosomal dominant polycystic kidney disease
 A:Reference number: A49367; MUID:94140377; PMID:8307582
 A:Accession: A49367
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-519 <WEI>
 A:Cross-references: GB:U02609; NID:9414535; PIDN:AAA18945.1; PID:9414536
 C:Genetics:
 A:Gene: sazd
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:77-110/Domain: WD repeat homology <WD1>
 F:258-291/Domain: WD repeat homology <WD2>
 F:361-394/Domain: WD repeat homology <WD3>

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 519;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 LFFSSATD 72
 |||||
 Db 53 LFFSSATD 60

RESULT 9
 G02075
 transcription repressor zinc finger protein 85 - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
 C:Accession: G02075
 R:Poncellet, D.A.
 A:Submitted to the EMBL Data Library, September 1995
 A:Reference number: G09169
 A:Accession: G02075
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-595 <PON>
 A:Cross-references: EMBL:U05376; NID:91017721; PIDN:AAA79179.1; PID:91017722
 C:Genetics:
 A:Gene: GDB:ZNF85
 A:Cross-references: GDB:133279
 A:Map position: 19p12-19p12
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 595;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
 |||||
 Db 50 PDLITCLE 57

RESULT 10
 D81337
 probable ATP/GTP-binding protein Cj0668 [Imported] - Campylobacter jejuni (strain NC
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: D81337
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanyilet, A.; Whitehead, S.; Bar
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: D81337
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB72945.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0668

Query Match
 Best Local Similarity 3.2%; Score 7; DB 2; Length 135;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 EGLLANG 201
 |||||
 Db 80 EGLLANG 86

RESULT 11
 JE0031
 nitrogenase (EC 1.18.6.1) molybdenum-iron protein nifX - Rhodobacter capsulatus
 C:Species: Rhodobacter capsulatus
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 19-Jan-2001
 C:Accession: JE0031; S04417
 R:Morono-Vivian, C.; Schmehl, M.; Masepohl, B.; Arnold, W.; Klipp, W.
 Mol. Gen. Genet. 216, 353-363, 1989
 A:Title: DNA sequence and genetic analysis of the Rhodobacter capsulatus nifX gene
 A:Reference number: JE0029; MUID:89313674; PMID:2747620
 A:Accession: JE0031
 A:Molecule type: DNA
 A:Residues: 1-159 <MOR>
 A:Cross-references: GB:X17433; NID:946085; PIDN:CAA3474.1; PID:946088
 C:Genetics:
 A:Gene: nifX
 C:Keywords: ATP; nitrogen fixation; oxidoreductase

Query Match
 Best Local Similarity 3.2%; Score 7; DB 2; Length 159;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FSSATDQ 73
 |||||
 Db 60 FSSATDQ 66

RESULT 12
 T34009
 hypothetical protein Y49G5A.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34009
 R:Becker, M.; Tin-Molam, A.M.; Yeakum, M.
 A:Submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid Y49G5A.
 A:Reference number: Z21457

A:Accession: T34009
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-182 <BEC>
 A:Cross-references: EMBL:AF125968; PIDD:AMD14760.1; GSPDB:GN00023; CESP:Y4965A.1
 A:Experimental source: strain Bristol N2; clone Y4965A
 C:Genetics:
 A:Gene: CESP:Y4965A.1
 A:Map position: 5
 A:introns: 44/3; 82/1; 102/3

Query Match 3.2%; Score 7; DB 2; Length 182;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 SKFCPAG 143
 DB 162 SKFCPAG 168

RESULT 13

AG2720
 riboflavin synthase alpha chain [imported] - Agrobacterium tumefaciens (strain C58, Dupc
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
 C:Accession: AG2720
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tso, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: A02577; PMID:11743193
 A:Accession: AG2720
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-203 <KUR>
 A:Cross-references: GB:AE008688; PIDD:AA142181.1; PTD:q17739571; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: ribc
 A:Map position: circular chromosome
 C:Superfamily: riboflavin synthase alpha chain

Query Match 3.2%; Score 7; DB 2; Length 203;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 DELGCHI 183
 DB 96 DELGCHI 102

RESULT 14

C97502
 riboflavin synthase alpha chain (A7236918) [imported] - Agrobacterium tumefaciens (stra
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: C97502
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: C97502
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-204 <KUR>
 A:Cross-references: GB:AB007869; PIDD:AAK86972.1; PTD:q15156208; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_2161
 A:Map position: circular chromosome

C:Superfamily: riboflavin synthase alpha chain

Query Match 3.2%; Score 7; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 DELGCHI 183
 DB 97 DELGCHI 103

RESULT 15

JF0074
 beta-lactamase (EC 3.5.2.6), cefotaxime-hydrolyzing - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 22-Apr-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
 C:Accession: JF0074; 141146
 R:ishii, Y.; Ohno, A.; Taguchi, H.; Imajo, S.; Ishiguro, M.; Matsuzawa, H.
 submitted to JIPID, July 1995
 A:Description: Cloning and the sequence analysis of a cefotaxime-hydrolyzing class A
 A:Reference number: JF0074
 A:Accession: JF0074
 A:Molecule type: protein
 A:Residues: 1-291 <ISR>
 R:ishii, Y.; Ohno, A.; Taguchi, H.; Matsuzawa, H.; Yamaguchi, K.
 Antimicrob. Agents Chemother. 39, 2269-2275, 1995
 A:Title: Cloning and sequence analysis of the gene for a cefotaxime-hydrolyzing class
 A:Reference number: 141146; MUID:96109425; PMID:8619581
 A:Accession: 141146
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-89, 'D', '91-291 <RES>
 A:Cross-references: GB:D37830; NID:q1435024; PIDD:BA07082.1; PTD:q1037162
 C:Superfamily: beta-lactamase I
 C:Keywords: hydrolase

Query Match 3.2%; Score 7; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSAT 71
 DB 19 LFFSAT 25

Search completed: May 15, 2003, 13:25:58
 Job time: 22.2527 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:17:05 ; Search time 6.53218 Seconds

(without alignments)
1390.548 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDCGCHIVTSQDSGTMT.....NGVLNRHGSLSKRFLETPP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 89080

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.7	195	1 ZN90_HUMAN	003938 homo sapien
2	8	3.7	353	1 Z254_HUMAN	075437 homo sapien
3	8	3.7	418	1 Z117_HUMAN	003924 homo sapien
4	8	3.7	439	1 UHPC_ECOLI	P09836 escherichia
5	8	3.7	519	1 TBL3_HUMAN	012788 homo sapien
6	8	3.7	535	1 Z257_HUMAN	092261 homo sapien
7	8	3.7	574	1 YET3_HUMAN	092255 homo sapien
8	8	3.7	595	1 ZN85_HUMAN	003923 homo sapien
9	7	3.2	159	1 NITX_RHOCA	P19078 rhodobacter
10	7	3.2	180	1 SYB_APICA	P35589 aplysia cal
11	7	3.2	291	1 BLC2_SALTY	P74841 salmonella
12	7	3.2	291	1 BLC3_SALTY	033807 salmonella
13	7	3.2	291	1 BLC5_SALTY	065975 salmonella
14	7	3.2	291	1 BLC6_SALTY	065976 salmonella
15	7	3.2	291	1 BLC7_SALTY	047066 escherichia
16	7	3.2	302	1 BCGH_RHOSE	092306 rhodobacter
17	7	3.2	359	1 SIA9_MOUSE	088829 mus musculu
18	7	3.2	384	1 PURR_BACSD	P39771 bacillus su
19	7	3.2	385	1 RECF_MYCTU	059586 mycobacteri
20	7	3.2	419	1 K1C4_XENLA	P05781 xenopus lae
21	7	3.2	494	1 VPE_CITSI	P49043 citrus sine
22	7	3.2	607	1 RECO_ECOLI	P15043 escherichia
23	7	3.2	642	1 YMG3_YEAST	P40308 saccharomyc
24	6	2.7	102	1 RPOL_AERPE	095621 aeropyrum p
25	6	2.7	110	1 KDFG_ERMCH	005527 aerwinia chr
26	6	2.7	122	1 RL18_BUCAL	P57575 buchnera ap
27	6	2.7	126	1 YIGF_ECOLI	P27842 escherichia
28	6	2.7	126	1 YIGF_SALTY	P31139 salmonella
29	6	2.7	137	1 ATPE_SYNP7	P07892 synecococc
30	6	2.7	138	1 ATPE_SYNP1	005375 synecococc
31	6	2.7	140	1 NT7_CYPCA	093474 cyprinus ca
32	6	2.7	141	1 V16K_TRVPS	P05075 tobacco rat
33	6	2.7	141	1 V16K_TRVSY	P05077 tobacco rat

34	6	2.7	141	1 V16K_TRVTC	P05076 tobacco rat
35	6	2.7	146	1 LU01_APICA	000676 aplysia cal
36	6	2.7	148	1 GVPJ_AKASP	089011 anabaena sp
37	6	2.7	148	1 NRDI_CORGL	09464 corynebacte
38	6	2.7	151	1 LCT2_HUMAN	014960 homo sapien
39	6	2.7	157	1 COAD_HELPY	026010 helicobacte
40	6	2.7	160	1 IF52_LYCES	09axq5 lycopersico
41	6	2.7	160	1 IF53_SOLTU	P56335 solanum tub
42	6	2.7	160	1 UBCC_SCHPO	09p611 schizosacch
43	6	2.7	170	1 OPR3_NEIMC	P10171 neisseria m
44	6	2.7	170	1 Y780_AOUAE	066975 aquifex aeo
45	6	2.7	173	1 NHP2_YEAST	P32495 saccharomyc

ALIGNMENTS

RESULT 1	ZN90_HUMAN	STANDARD;	PRT;	195 AA.
AC	003938;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Zinc finger protein 90 (Zinc finger protein HTF9) (Fragment).			
GN	ZNFP90.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91219421; PubMed=2023909;			
RA	Balliford E.J., Poncelet D.A., Lecocq P.J., Revelant O.,			
RA	Marital J.A.,			
RT	"The evolutionarily conserved Kruppel-associated box domain defines a			
RT	subfamily of eukaryotic multifingered proteins."			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).			
CC	- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.			
CC	- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.			
CC	- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.			
CC	- SIMILARITY: CONTAINS 1 KRAB DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL: M61870; AAA36028.1; -			
DR	PIR: E3384; E39384.			
DR	Genew; HGNC:13165; ZNF90.			
DR	MIM; 603973; -			
DR	InterPro: IPR001909; KRAB.			
DR	InterPro: IPR000822; ZnfC2H2.			
DR	Pfam; PF00096; zf-C2H2; 1.			
DR	Pfam; PF01352; KRAB; 1.			
DR	SMART; SM00349; KRAB; 1.			
DR	SMART; SM00355; ZNF_C2H2; 1.			
DR	PROSITE; PS00805; KRAB; 1.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.			
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.			
KW	Transcription regulation; Zinc-finger; DNA-binding; Metal-binding; Nuclear protein; Repeat.			
KW	DOMAIN			
FT	4			
FT	ZN_FING	145	167	KRAB.
FT	ZN_FING	173	195	C2H2-TYPE (DEGENERATE).
FT	NON_TER	195	195	C2H2-TYPE.
SQ	SEQUENCE	195 AA;	22562 MW;	0E808A1A62342608 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 50 PDLITCLE 57

RESULT 2
2254_HUMAN
ID 2254_HUMAN STANDARD; PRT; 353 AA.
AC 075437;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Zinc finger protein 254 (Bone marrow zinc finger 5) (BMZF-5)
GN (hematopoietic cell derived zinc finger protein 1) (HD-ZNF1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mo M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
Wang Y.-X., Chen S.-J., Chen Z.;
RT "Identification of genes expressed in human CD34(+) hematopoietic
stem/progenitor cells by expressed sequence tags and efficient full-
length cDNA cloning."
RT Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=20054457; PubMed=10585455;
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
Shi S.-L., Zhou J., Fu G., Mo M., Chen S.-J., Yu L., Chen Z.;
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
hematopoietic cells and identification of a novel transregulatory
domain KRNH."
RT J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nucleus (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; AF054180; AAC39913.1; -
CC Genew; HGNC:13047; ZNF254.
CC MIM: 604768; -
CC InterPro: IPR001909; KRAB.
CC InterPro: IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf_C2H2; 4.
CC PRINTS; PR00048; ZNCFINGER.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00349; KRAB; 1.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00805; KRAB; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
CC Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

Query Match 3.7%; Score 8; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 59 PDLITCLE 66

RESULT 3
2117_HUMAN
ID 2117_HUMAN STANDARD; PRT; 418 AA.
AC 003924;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 117 (Zinc finger protein HPF9) (Fragment).
GN ZNF117.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 1;
RP SEQUENCE OF 4-418 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=69377476; PubMed=2505992;
RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,
Belayew A., Martial J.A.;
RT "The human genome contains hundreds of genes coding for finger
proteins of the Kruppel type."
RT DNA 8:377-387(1989).
RN 1;
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
Martial J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
subfamily of eukaryotic multifingered proteins."
RT Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nucleus (Probable).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; M27879; -; NOT ANNOTATED_CDS.
CC EMBL; M61867; AA58666.1; -
CC PIR; B39384; B39384.
CC HSP; P03001; ITF6.
CC Genew; HGNC:12897; ZNF117.
CC MIM; 194624; -
CC InterPro: IPR001909; KRAB.
CC InterPro: IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf_C2H2; 10.
CC ProDom; PD000003; Znf_C2H2; 5.

DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; ZNF_C2H2; 10.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
 DR Transcription regulation; zinc-finger; DNA-binding; Metal-binding;
 KW Nucleic acid; protein; Repeat.
 FT DOMAIN 1 46 KRAB.
 FT ZN_FING 116 417 ZINC_FINGERS.
 FT ZN_FING 116 138 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 144 166 C2H2-TYPE.
 FT ZN_FING 172 194 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 200 222 C2H2-TYPE.
 FT ZN_FING 228 250 C2H2-TYPE.
 FT ZN_FING 256 278 C2H2-TYPE.
 FT ZN_FING 284 306 C2H2-TYPE.
 FT ZN_FING 312 334 C2H2-TYPE.
 FT ZN_FING 340 362 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 368 390 C2H2-TYPE.
 FT ZN_FING 396 417 C2H2-TYPE.
 FT CONFLICT 75 75 R -> G (IN REF. 2).
 FT CONFLICT 78 79 GY -> RH (IN REF. 2).
 SQ SEQUENCE 418 AA; 48941 MW; CDB554CB69EB0869 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDITCIE 126
 DB 21 PDITCIE 28
 RESULT 4
 UHPC_ECOLI STANDARD; PRT; 439 AA.
 ID UHPC_ECOLI
 AC P09836; P76728;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Regulatory protein uhpc.
 GN UHPC OR B3667.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87279903; Pubmed-3301805;
 RA Friedrich M.J.; Kadner R.J.;
 RT "Nucleotide sequence of the uhpc region of Escherichia coli.";
 RL J. Bacteriol. 169:3556-3563(1987).
 RN [2]
 RP REVISIONS.
 RA Kadner R.J.;
 RL Submitted (Aug-1990) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92234930; Pubmed-1569007;
 RA Island M.D.; Wei B.-Y.; Kadner R.J.;
 RT "Structure and function of the uhpc genes for the sugar phosphate
 transport system in Escherichia coli and Salmonella typhimurium.";
 RL J. Bacteriol. 174:2754-2762(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-93351513; Pubmed-7686882;
 RA Burland V.D.; Plunkett G. II; Daniels D.L.; Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
 genome: organizational symmetry around the origin of replication.";
 RN Genomics 16:351-361(1993).
 RN [5]

RP POSSIBLE ROLE.
 RX MEDLINE-93352407; Pubmed-8349544;
 RA Island M.D.; Kadner R.J.;
 RT "Interplay between the membrane-associated Uhpb and Uhpc regulatory
 proteins.";
 RL J. Bacteriol. 175:5028-5034(1993).
 CC -1- FUNCTION: UHPC IS REQUIRED FOR UHPB EXPRESSION, IT MAY ACT JOINTLY
 WITH THE UHPB SIGNALING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M17102; AAA24722.1; ALT_INIT.
 DR EMBL; M89479; AAA24726.1; -.
 DR EMBL; I10328; AAA62019.1; ALT_INIT.
 DR EMBL; AE000444; AAC76690.1; ALT_INIT.
 DR PIR; G41853; KEGCUC.
 DR EcoGene; EG11053; uhpc.
 DR InterPro; IPR000849; GIPr transporter.
 DR InterPro; IPR003662; sub transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR TIGRfam; TIGR00881; ZAO104; 1.
 DR PROSITE; PS00942; GLPT; 1.
 DR KW Transport; Sugar transport; Transmembrane; Inner membrane;
 KM Complete proteome.
 FT TRANSMEM 67 87 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 163 183 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 245 265 POTENTIAL.
 FT TRANSMEM 290 310 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 411 431 POTENTIAL.
 SQ SEQUENCE 439 AA; 48256 MW; D8097864B77700 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 439;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 198 LANGVLSR 205
 DB 57 LANGVLSR 64
 RESULT 5
 TBL3_HUMAN STANDARD; PRT; 519 AA.
 ID TBL3_HUMAN
 AC Q12788;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE WD-repeat protein SAZD (Transducin beta-like 3 protein).
 GN TBL3 OR SAZD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94140377; Pubmed-8307582;
 RA Weinstein Saslow D.L.; Germino G.G.; Somlo S.; Reiders S.T.;
 RT "A transducin-like gene maps to the autosomal dominant polycystic
 kidney disease gene region.";
 RL Genomics 18:709-711(1993).

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U02609; AAA18945.1; -.

DR Genew; HGNC:11587; TBL3.

DR MIM; 605915; -.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40.11.

DR PRINTS; PR00320; GPROTEINRPT.

DR PRODOM; PD000018; WD40.2.

DR SMART; SM00320; WD40.9.

DR PROSITE; PS00678; WD_REPEATS_1; 2.

DR PROSITE; PS50082; WD_REPEATS_2; 7.

DR PROSITE; PS50294; WD_REPEATS_REGION; 2.

KW Repeat; WD repeat.

FT REPEAT 1 23 WD 1.

FT REPEAT 35 67 WD 2.

FT REPEAT 79 109 WD 3.

FT REPEAT 260 290 WD 4.

FT REPEAT 305 336 WD 5.

FT REPEAT 363 393 WD 6.

FT REPEAT 447 477 WD 7.

SO SEQUENCE 519 AA; 56047 MW; EB08B46952499BD CRC64;

Query Match 3.7%; Score 8; DB 1; Length 519;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSATD 72

Db 53 LFFSSATD 60

RESULT 6

ID Z237_HUMAN STANDARD; PRT; 535 AA.

AC Q9Y2Q1;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).

GN ZNF257 OR BMZF4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=20054457; PubMed=10585455;

RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,

RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;

RT "Molecular cloning of six novel Kruppel-like zinc finger genes from

RT hematopoietic cells and identification of a novel transregulatory

RT domain KRMB.";

RL J. Biol. Chem. 274:35741-35748(1999).

CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-

CC FINGER PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF070651; AAD20957.1; -.

DR HSSP; P08047; ISP2.

DR Genew; HGNC:13498; ZNF257.

DR MIM; 606957; -.

DR InterPro; IPR001909; KRAB.

DR InterPro; IPR000822; ZnF_C2H2.

DR Pfam; PF00096; Zf-C2H2; 10.

DR Pfam; PF01352; KRAB; 1.

DR PRINTS; PR00048; ZINC_FINGER.

DR PRODOM; PD000003; ZnF_C2H2; 9.

DR SMART; SM00349; KRAB; 1.

DR SMART; SM00355; ZnF_C2H2; 10.

DR PROSITE; PS50805; KRAB; 1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.

KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 4 75 KRAB.

FT ZN_FING 173 509 ZINC_FINGERS.

FT ZN_FING 173 195 C2H2-TYPE.

FT ZN_FING 201 223 C2H2-TYPE.

FT ZN_FING 229 251 C2H2-TYPE.

FT ZN_FING 257 282 C2H2-TYPE (DEGENERATE).

FT ZN_FING 288 310 C2H2-TYPE.

FT ZN_FING 316 338 C2H2-TYPE.

FT ZN_FING 344 366 C2H2-TYPE.

FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).

FT ZN_FING 403 425 C2H2-TYPE.

FT ZN_FING 431 453 C2H2-TYPE.

FT ZN_FING 459 481 C2H2-TYPE.

FT ZN_FING 487 509 C2H2-TYPE.

SO SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 535;

Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

Db 50 PDLITCLE 57

RESULT 7

ID YE73_HUMAN STANDARD; PRT; 574 AA.

AC Q9P255;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical zinc finger protein KIA1473 (fragment).

GN KIA1473.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=20277482; PubMed=10819331;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XVIII.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.";

RL DNA Res. 7:143-150(2000).

CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-

CC FINGER PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB040906; BAA9597.1; -
 DR HSSP: P07248; 1ARD.
 DR InterPro: IPR001909; KRAB.
 DR Pfam: PF00096; zf_C2H2_13.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 13.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
 KM Hypothetical protein; Transcription regulation; DNA-binding;
 KM Zinc-finger; Metal-binding; Nuclear protein; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 15 86 KRAB.
 FT ZN_FING 184 542 ZINC_FINGERS.
 FT ZN_FING 184 206 C2H2-TYPE.
 FT ZN_FING 212 234 C2H2-TYPE.
 FT ZN_FING 240 262 C2H2-TYPE.
 FT ZN_FING 268 290 C2H2-TYPE.
 FT ZN_FING 296 318 C2H2-TYPE.
 FT ZN_FING 324 346 C2H2-TYPE.
 FT ZN_FING 352 374 C2H2-TYPE.
 FT ZN_FING 380 402 C2H2-TYPE.
 FT ZN_FING 408 430 C2H2-TYPE.
 FT ZN_FING 436 458 C2H2-TYPE.
 FT ZN_FING 464 486 C2H2-TYPE.
 FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 520 542 C2H2-TYPE.
 SO SEQUENCE 574 AA; 65951 MM; 48255F1176ECAF50 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 574;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDITCIE 126
 Db 61 PDITCIE 68
 RESULT 8
 ZNF85_HUMAN STANDARD; PRT; 595 AA.
 ID ZNF85_HUMAN
 AC 003923;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Zinc finger protein 85 (Zinc finger protein HPP4) (HPP1).
 GN ZNF85.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Placenta;
 RX MEDLINE=9905357; PubMed=9839802;
 RA Marcel D.A., Belleville E.J., Bastiaens P.V., Demolite M.A.,
 RA Marne J.C., Pendergast H., Alami Y., Devos N., Lecocq P.J., Ogawa T.,
 RA Muller M., Marital J.A.,
 RT Functional analysis of ZNF85 KRAB zinc finger protein, a member of
 RT the highly homologous ZNF91 family.";

 RL DNA Cell Biol. 17:931-943(1998).
 RN [2]
 RP SEQUENCE OF 1-196 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Belleville E.J., Ponder D.A., Lecocq P.J., Revelant O.,
 RA Marital J.A.,
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins."
 RT Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN TESTICULAR
 CC TISSUES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U35376; AAA79179.1; -
 DR EMBL: M61866; AAA52689.1; -
 DR EMBL: M61868; AAA58671.1; -
 DR PIR: A39384; A39384.
 DR PIR: C39384; C39384.
 DR HSSP: P08048; ZNF.
 DR TRANSFAC: T04990; -
 DR Genev: HGNC:13160; ZNF85.
 DR MIM: 603899; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf_C2H2; 15.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; Znf_C2H2; 13.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 15.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
 KM Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
 KM Nuclear protein; Repeat; Repressor.
 FT DOMAIN 4 75 KRAB.
 FT ZN_FING 146 588 ZINC_FINGERS.
 FT ZN_FING 146 168 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 174 196 C2H2-TYPE.
 FT ZN_FING 202 224 C2H2-TYPE.
 FT ZN_FING 230 252 C2H2-TYPE.
 FT ZN_FING 258 280 C2H2-TYPE.
 FT ZN_FING 286 308 C2H2-TYPE.
 FT ZN_FING 314 336 C2H2-TYPE.
 FT ZN_FING 342 364 C2H2-TYPE.
 FT ZN_FING 370 392 C2H2-TYPE.
 FT ZN_FING 398 420 C2H2-TYPE.
 FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 454 476 C2H2-TYPE.
 FT ZN_FING 482 504 C2H2-TYPE.
 FT ZN_FING 510 532 C2H2-TYPE.
 FT ZN_FING 538 560 C2H2-TYPE.
 FT ZN_FING 566 588 C2H2-TYPE.
 FT ZN_FING 584 606 C2H2-TYPE.
 FT CONFLICT 84 84 R -> Q (IN REF. 2; AAA52689).
 FT CONFLICT 115 115 T -> I (IN REF. 2; AAA58671).
 FT CONFLICT 177 177 T -> R (IN REF. 2; AAA58671).
 FT CONFLICT 184 184 G -> R (IN REF. 2; AAA58671).
 SO SEQUENCE 595 AA; 68718 MM; 44AA0A236D243B CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 595;

Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITTLE 126
|111111|
DB 50 PDLITTLE 57

RESULT 9

NIFX_RHOCA
ID NIFX_RHOCA STANDARD; PRT; 159 AA.
AC P19078;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Nifx protein.
GN NIFX.
OS Rhodobacter capsulatus (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
CC Rhodobacter.
CC NCB1_TaxID=1061;
CC [1]
CC SEQUENCE FROM N.A.

RA MEDLINE=89313674; PubMed=2747620;
RA "Moreno-Vivian C., Schmehl M., Masepohl B., Arnold W., Klipp W.,
RT "DNA sequence and genetic analysis of the Rhodobacter capsulatus
RT alienX gene region: homology between Nifx and NifB suggests
RT involvement of Nifx in processing of the iron-molybdenum cofactor."
RL Mol. Gen. Genet. 216:353-363(1989).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE PROCESSING OF THE IRON-
CC MOLYBDENUM COFACTOR.
CC -1- SIMILARITY: TO NIFB.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; X17433; CAA5474.1; -
DR PIR; J00031; J00031.
DR InterPro: IPR003731; DUF153.
DR Pfam; PF02579; DUF153; 1.
RW Nitrogen fixation.
SQ SEQUENCE 159 AA; 17456 MW; 8470D344B6C71005 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FSSARDQ 73
|111111|
DB 60 FSSARDQ 66

RESULT 10

SYB_APLCA STANDARD; PRT; 180 AA.
AC P35589;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Synaptoleptin.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
CC Aplysiidae; Aplysia.
CC NCB1_TaxID=6500;
CC [1]

RN SEQUENCE FROM N.A.
RP TISSUE=Ganglion;
RX MEDLINE=94255392; PubMed=8197120;

RA Yamasaki S., Hu Y., Binz T., Kalkuhl A., Kurazono H., Tamura T.,
RA Jahn R., Kandel E., Niemann H.;
RT "Synaptoleptin/vesicle-associated membrane protein (VAMP) of Aplysia
RT californica: structure and proteolysis by tetanus toxin and botulinum
RT neurotoxins type D and F."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4688-4692(1994).
CC -1- FUNCTION: INTRINSIC MEMBRANE PROTEIN OF SMALL SYNAPTIC
CC VESICLES.

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NEURONAL SYNAPTIC
CC VESICLES.
CC -1- SIMILARITY: BELONGS TO THE SYNAPTOLEPTIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; U00997; AAA19442.1; -
DR InterPro: IPR001388; Synaptoleptin.
DR Pfam; PF00957; synaptoleptin; 1.
DR PRINTS; PR00219; SYNAPTOLEPTIN.
DR PRODOM; PD001229; synaptoleptin; 1.
DR PROSITE; PS00417; SYNAPTOLEPTIN; 1.
KW Synapse; Transmembrane; Synaptoleptin; Signal-anchor.
FT DOMAIN 1 84 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 85 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 105 180 VESICULAR (PROBABLE).
SQ SEQUENCE 180 AA; 19746 MW; ADC780DA74CF347 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VTSQDSG 17
|111111|
DB 105 VTSQDSG 111

RESULT 11
BLC2_SALTY
ID BLC2_SALTY STANDARD; PRT; 291 AA.
AC P74841;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-lactamase CTX-M-2 precursor (EC 3.5.2.6) (Cefotaximase 2).
GN BLa.

OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Bacteria.
CC Salmonella.
CC NCB1_TaxID=602;
CC [1]
CC SEQUENCE FROM N.A.

RN STRAIN=CAS-5;
RC MEDLINE=96431831; PubMed=8834913;
RA Bauernfeind A., Stemplinger I., Jungwirth R., Casellas J.M.;
RT "Sequences of beta-lactamase genes encoding CTX-M-1 (MEN-1) and CTX-M-
RT 2 and relationship of their amino acid sequences with those of other
RT beta-lactamases."
RL Antimicrob. Agents Chemother. 40:509-513(1996).
CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
CC amino acid.

CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).

DR EMBL: X92507; CA63263.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-2.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31377 MW; 82277CD9CDA1E7C0 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 12
BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 03807;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-lactamase CTX-M-3 precursor (EC 3.5.2.6) (cefotaxime 3).
GN BLA.
OS Salmonella typhimurium.
OC Plasmid PMS1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN (1)
RP SEQUENCE FROM N.A.
RA Gazouli M.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O - a substituted beta-amino acid.
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).

DR EMBL: Y14156; CA74573.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-3.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31254 MW; 2E22E251008DF7C6 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 13
BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 065976;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-lactamase CTX-M-5 precursor (EC 3.5.2.6) (cefotaxime 5).
GN BLA.
OS Salmonella typhimurium.
OC Plasmid PAS30.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN (1)
RP SEQUENCE FROM N.A.
RA Gazouli M.; Tzelepi E.; Markogiannakis A.; Legakis N.; Tzouveleakis L.;
RT "Two novel plasmid-mediated cefotaxime-hydrolyzing beta-lactamases
RT (CTX-M-5 and CTX-M-6) from Salmonella typhimurium."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O - a substituted beta-amino acid.
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).

DR EMBL: AJ005044; CA06311.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-5.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31251 MW; DCC1C3F73789894E CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 14
BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 065976;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Beta-lactamase CTX-M-6 precursor (EC 3.5.2.6) (cefotaximase 6).
 GN BIA.
 OS *Salmonella typhimurium*.
 CC Plasmid PAS31.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC *Salmonella*.
 NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AS31;
 RA Gazouli M., Tzelepi E., Markogiannakis A., Legakis N., Tzouveleakis L.;
 RT "Two novel plasmid-mediated cefotaxime-hydrolyzing beta-lactamases
 (CTX-M-5 and CTX-M-6) from *Salmonella typhimurium*.";
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBD databases.
 CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
 CC amino acid.
 CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ005045; CA06312.1; -
 DR HSSP: Q47066; IBZA.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
 DR KMW: Hydrolyase; Antibiotic resistance; Plasmid; Signal.
 FT SIGNAL 1 28 BY SIMILARITY.
 FT CHAIN 29 291 BETA-LACTAMASE CTX-M-6.
 FT ACT_SITE 73 73 BY SIMILARITY.
 FT BINDING 237 239 SUBSTRATE (BT SIMILARITY).
 SQ SEQUENCE 291 AA; 31206 MW; DC95EAF677531053 CRC64;
 Query Match 3.2%; Score 7; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 LFFSSAT 71
 Db 19 LFFSSAT 25
 RESULT 15
 BLT1_ECOLI STANDARD; PRT; 291 AA.
 AC Q47066;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Beta-lactamase Toho-1 precursor (EC 3.5.2.6).
 GN BIA.
 OS *Escherichia coli*.
 CC Plasmid.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC *Escherichia*.
 NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THH12191;
 RX MEDLINE=96109425; PubMed=8619581;
 RA Ishii Y., Ohno A., Taguchi H., Imajo S., Ishiguro M., Matsuzawa H.;
 RT "Cloning and sequence of the gene encoding a cefotaxime-hydrolyzing
 RT class A beta-lactamase isolated from *Escherichia coli*.";
 RL Antimicrob. Agents Chemother. 39:2269-2275(1995).
 RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT ALA-169.
 RC STRAIN=THH12191;
 RX MEDLINE=99126537; PubMed=9925786;
 RA Iibuka A., Taguchi A., Ishiguro M., Fushinobu S., Ishii Y.,
 RA Kamitori S., Okuyama K., Yamaguchi K., Kono M., Matsuzawa H.;
 RT "Crystal structure of the E166 mutant of extended-spectrum
 RT beta-lactamase toho-1 at 1.8 Å resolution".
 RL J. Mol. Biol. 285:2079-2087(1999).
 CC -1- FUNCTION: HAS STRONG CEFOTAXIME-HYDROLYZING ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D37830; BAA07082.1; -
 DR PDB: 1BZA; 27-Apr-99.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
 DR KMW: Hydrolyase; Antibiotic resistance; Signal; Plasmid; 3D-structure.
 FT SIGNAL 1 29 BY SIMILARITY.
 FT CHAIN 30 291 BETA-LACTAMASE TOHO-1.
 FT ACT_SITE 73 73 SUBSTRATE.
 FT BINDING 237 239
 SQ SEQUENCE 291 AA; 31447 MW; 83FC0CD9CD41E7C0 CRC64;
 Query Match 3.2%; Score 7; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 LFFSSAT 71
 Db 19 LFFSSAT 25

Search completed: May 15, 2003, 13:23:08
 Job time : 9.53218 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:18:56 ; Search time 18.9089 Seconds
(without alignments)
2386.400 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDGCGHIVTSODSGTMT.....NCVLSRHGSLSEKRELFPTTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 504360

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	100.0	503	11	09D4J3 mus musculu
2	201	91.8	460	11	09D696 mus musculu
3	145	66.2	432	11	08R327 mus musculu
4	8	3.7	195	17	08TJ73 methanosarc
5	8	3.7	216	4	096H40 homo sapien
6	8	3.7	336	10	09MAA9 Oryzae arabidopsis
7	8	3.7	364	4	09H8G1 homo sapien
8	8	3.7	440	16	08X543 Oryzae escherichia
9	8	3.7	470	4	09B2E6 homo sapien
10	8	3.7	519	4	096H04 homo sapien
11	8	3.7	524	4	0965C4 homo sapien
12	8	3.7	536	4	096B85 homo sapien
13	7	3.2	132	5	08T1H2 dictyostell
14	7	3.2	135	16	09PPN3 campylobact
15	7	3.2	156	17	08TKH7 methanosarc
16	7	3.2	182	5	09DMA2 caenorhabd

17	7	3.2	194	11	091YF2 mus musculu
18	7	3.2	204	16	08UG72 agrobacteri
19	7	3.2	205	16	0983B1 rhizobium i
20	7	3.2	206	16	0920U4 rhizobium m
21	7	3.2	251	7	031296 sciurus abe
22	7	3.2	272	16	08R2M5 Oryzae ralsonia s
23	7	3.2	276	2	08VQ11 acinetobact
24	7	3.2	291	2	0936V5 proteus mir
25	7	3.2	291	2	09RLX4 klyvera as
26	7	3.2	291	2	08VN25 klyvera as
27	7	3.2	291	2	08VN24 klyvera as
28	7	3.2	291	2	08VN23 klyvera as
29	7	3.2	291	2	08VN22 klyvera as
30	7	3.2	291	2	08VKT5 klyvera as
31	7	3.2	291	2	034177 salmoneella
32	7	3.2	294	2	0931I6 burkholderi
33	7	3.2	300	5	08SU29 encephalito
34	7	3.2	314	17	097Z87 sulfobolus
35	7	3.2	320	5	09XMB9 caenorhabd
36	7	3.2	332	13	091333 rana sylvat
37	7	3.2	350	4	09NV85 homo sapien
38	7	3.2	350	4	09H813 homo sapien
39	7	3.2	350	11	09CZV9 mus musculu
40	7	3.2	350	11	09D771 mus musculu
41	7	3.2	359	11	09QWF8 mus musculu
42	7	3.2	359	11	088829 mus musculu
43	7	3.2	368	5	09GY66 leishmania
44	7	3.2	381	10	09FNP5 arabidopsis
45	7	3.2	385	3	08X058 neurospora

ALIGNMENTS

RESULT 1	ID	09D4J3	PRELIMINARY:	PRT:	503 AA.
AC	09D4J3	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)				
DE	4631413K1R1K protein.				
GN	4631413K1R1K.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;				
RX	MEDLINE=21085660; PubMed=11217651;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamana K.,				
RA	Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Glass C., King B., Kociba H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,				
RA	Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Williams L.,				
RA	Yuzhaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,				
RA	Hayashizaki Y.,				
RT	*Functional annotation of a full-length mouse cDNA collection.*;				
RL	Nature 409:685-690(2001).				
CC	-1- SIMILARITY: CONTAINS 1 CDB DOMAIN.				
DR	EMBL: AK016485; BAB30265.1; -.				

DR MGD; MGI:1913936; 4631413K11R1K.
 DR InterPro: IPR000859; CUB_dom.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SQ SEQUENCE 503 AA; 54547 MW; FEB121E845CA06B CRC64;

Query Match 100.0%; Score 219; DB 11; Length 503;
 Best Local Similarity 100.0%; Pred. No. 1.2e-224;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELDDGCHITSDSGMTSKNPGTPNTYVCEKITVYKGRKLLRLGDLNIESKTC 60
 |||||||
 DB 26 EELDDGCHITSDSGMTSKNPGTPNTYVCEKITVYKGRKLLRLGDLNIESKTC 85

QY 61 ASDYLFSATDQGYPCGSAVPEKRLNSENVTYLFKSGSHISGRGFLTYASSDHPD 120
 |||||||
 DB 86 ASDYLFSATDQGYPCGSAVPEKRLNSENVTYLFKSGSHISGRGFLTYASSDHPD 145

QY 121 LITCLERSHYEERKSKFCPCAGCDIADISGNKDGYPDTSILCKAIIHAGITDELG 180
 |||||||
 DB 146 LITCLERSHYEERKSKFCPCAGCDIADISGNKDGYPDTSILCKAIIHAGITDELG 205

QY 181 GHINLQSKGSHYEGGLANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 206 GHINLQSKGSHYEGGLANGVLSRHGSLSEKRFLETPP 244

RESULT 2
 Q9D696 PRELIMINARY; PRT; 460 AA.

AC Q9D696;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DE 4631413K11R1K protein.
 GN 4631413K11R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 Schiraldi L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Orido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 Hayashizaki Y.,
 RA Functional annotation of a full-length mouse cDNA collection.;
 RT Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK014521; BAB29409.1; -;
 DR MGD; MGI:1913936; 4631413K11R1K.
 DR InterPro: IPR000859; CUB_domain.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR SMART: SM00042; CUB; 1.

DR PROSITE: PS01180; CUB; 1.
 SQ SEQUENCE 460 AA; 50334 MW; 3CF3356F70A8B93 CRC64;

Query Match 91.8%; Score 201; DB 11; Length 460;
 Best Local Similarity 100.0%; Pred. No. 1.8e-205;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MTSKNYPTDPTNYVCEKITVYKGRKLLRLGDLNIESKTCASDYLLFSSATDQGYPC 78
 |||||||
 DB 1 MTSKNYPTDPTNYVCEKITVYKGRKLLRLGDLNIESKTCASDYLLFSSATDQGYPC 60

QY 79 GSMVAPKRLNSENVTYLFKSGSHISGRGFLTYASSDHPDITCLERSHYEERKSK 138
 |||||||
 DB 61 GSMVAPKRLNSENVTYLFKSGSHISGRGFLTYASSDHPDITCLERSHYEERKSK 120

QY 139 FCPAGCDIADISGNKDGYPDTSILCKAIIHAGITDELGHINLQSKGSHYEGGL 198
 |||||||
 DB 121 FCPAGCDIADISGNKDGYPDTSILCKAIIHAGITDELGHINLQSKGSHYEGGL 180

QY 199 ANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 181 ANGVLSRHGSLSEKRFLETPP 201

RESULT 3
 Q8R327 PRELIMINARY; PRT; 432 AA.

AC Q8R327;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Similar to RIKEN cDNA 4631413K11 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC026771; AAH26771.1; -;
 SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 66.2%; Score 145; DB 11; Length 432;
 Best Local Similarity 100.0%; Pred. No. 8.5e-146;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 GPGGSAVPEKRLNSENVTYLFKSGSHISGRGFLTYASSDHPDITCLERSHYEE 134
 |||||||
 DB 29 GPGGSAVPEKRLNSENVTYLFKSGSHISGRGFLTYASSDHPDITCLERSHYEE 88

QY 135 KYSKFCPCAGCDIADISGNKDGYPDTSILCKAIIHAGITDELGHINLQSKGSHY 194
 |||||||
 DB 89 KYSKFCPCAGCDIADISGNKDGYPDTSILCKAIIHAGITDELGHINLQSKGSHY 148

QY 195 EGLANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 149 EGLANGVLSRHGSLSEKRFLETPP 173

RESULT 4
 Q8UTJ3 PRELIMINARY; PRT; 195 AA.

AC Q8UTJ3;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein MA3694.
 GN MA3694.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_Taxid=2214;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CZA / ATCC 35395 / DSM 2834;
 RX MEDLINE-21929760; PubMed-11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., DeArrellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of *Methanocaldococcus jannaschii* reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AF011078; AAM07049.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 195 AA; 22019 MW; 909E97DB7359DC67 CRC64;

 Query Match 3.7%; Score 8; DB 17; Length 195;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 172 AGITDEL 179
 Db 98 AGITDEL 105

 RESULT 5
 ID 096H40 PRELIMINARY; PRT; 216 AA.
 AC 096H40;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-MAR-2002 (TEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
 DE Similar to DNA-binding protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LUNG;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC008936; AAH08936.1; -;
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF01352; KRAB.1.
 DR PROSITE; PSS0805; KRAB.1.
 DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 2.
 DR DNA-binding; Zinc-finger.
 KW SEQUENCE 216 AA; 25227 MW; B5C6CF182AFCEP2 CRC64;

 Query Match 3.7%; Score 8; DB 4; Length 216;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 119 PDLITCLE 126
 Db 50 PDLITCLE 57

 RESULT 6
 ID 09MAA9 PRELIMINARY; PRT; 336 AA.
 AC 09MAA9;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE T12H1.6 protein (Hypothetical 37.7 kDa protein).
 GN T12H1.6 OR AT3G05100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Beutlo M., Creasy T.H., Haas B., Wu D.,
 RA Rongning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At3g05100 (GI:15229893)."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Narusaka M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC009177; AAR27016.1; -;
 DR EMBL; AY070370; AA149868.1; -;
 DR EMBL; AY091328; AA14267.1; -;
 DR InterPro; IPR000051; SAM_bind.
 KW Hypothetical protein.
 SQ SEQUENCE 336 AA; 37719 MW; 6FA910F7B4B2FA85 CRC64;

 Query Match 3.7%; Score 8; DB 10; Length 336;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 189 KGISHYEG 196
 Db 121 KGISHYEG 128

 RESULT 7
 ID 09H8G1 PRELIMINARY; PRT; 364 AA.
 AC 09H8G1;
 DT 01-MAR-2001 (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE CDNA FLJ13659 fls. clone PLACE1011576, moderately similar to human
 DE kruppel related zinc finger protein (KRF10) mRNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RA Tsogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sato H., Sugano S., Shitatori A., Sudo H.,

RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakematsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Niimura K., Iwayanagi T.,
 RT "NDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK023721; BAB14656.1; -
 DR InterPro: IPR001909; KRA6.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF01352; KRA6; 1.
 DR Pfam: PF00096; Zf_C2H2; 3.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00349; KRA6; 1.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00805; KRA6; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SO SEQUENCE 364 AA; 42472 MW; 5043DA4C575842B0 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 364;
 Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 82 PDLITCLE 89

RESULT 8
 08X543 PRELIMINARY; PRT; 440 AA.

AC 08X543;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Regulator of uhpF.
 GN UHPC OR Z5157 OR ECS4604.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Pena N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobleck E.J., Davis N.W., Lim A., Dimmalanta E.T., Polamakis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).

DR EMBL: AE005598; AAG58866.1; -
 DR EMBL: AP002366; BAB38027.1; -
 DR InterPro: IPR000849; G1PT_transporter.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 DR TIGRfams: TIGR00881; 2A0104; 1.
 DR PROSITE: PS00942; G1PT; 1.

KW Complete proteome.
 SO SEQUENCE 440 AA; 48345 MW; C3442BAC11132448 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 440;
 Best Local Similarity 100.0%; Pred. No. 10;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 198 LANGVLSR 205
 |||||
 DB 58 LANGVLSR 65

RESULT 9
 09BZ6 PRELIMINARY; PRT; 470 AA.

AC 09BZ6;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE KRA6 zinc finger protein HZF26 (Fragment).
 GN HZF26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95169271; PubMed=7865130;
 RA Abrik M., Aveskogh M., Hellman L.;
 RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
 proteins expressed in the human monoblast cell line U-937.";
 RL DNA Cell Biol. 14:125-136(1995).
 [2]

RP SEQUENCE FROM N.A.
 RA Abrik M., Aveskogh M., Hellman L.;
 RL Submitted (Nov-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AF325191; AAK01422.1; -
 DR HSSP: P08046; 1AIG.
 DR InterPro: IPR001909; KRA6.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF01352; KRA6; 1.
 DR Pfam: PF00096; Zf_C2H2; 10.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR SMART: SM00349; KRA6; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS00805; KRA6; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 11.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 FT NON_TER 470
 SQ SEQUENCE 470 AA; 54961 MW; 997CB1C95D37E983 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 470;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 50 PDLITCLE 57

RESULT 10

096M04 PRELIMINARY; PRT; 519 AA.

AC 096M04;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE cDNA FLJ32933 fls, clone TEST12007466, moderately similar to zinc
 DE finger protein 91.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Houta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato K., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Nagatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isoegal T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK057495; BAB71510.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000832; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 12.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 519 AA; 60350 MW; 7E4D17E53EC7732C CRC64;

Query Match 3.7%; Score 8; DB 4; Length 519;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 59 PDLITCLE 66

RESULT 11
 ID 096JC4 PRELIMINARY; PRT; 524 AA.
 AC 096JC4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE KRAB zinc finger protein.
 GN KRI9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RX MEDLINE-21303228; PubMed-11410164;
 RA Mark C., Looman C., Ahrink M., Hellman L.;
 RT "Molecular cloning and preliminary functional analysis of two novel
 RT human KRAB zinc finger proteins, Hkr18 and Hkr19."
 RL DNA Cell Biol. 20:275-286(2001).
 DR EMBL: AF277624; AAK61307.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000832; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 12.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_11.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 524 AA; 60598 MW; F1AD4929DC67D105 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 524;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 62 PDLITCLE 69

RESULT 12
 ID 096BB5 PRELIMINARY; PRT; 536 AA.
 AC 096BB5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical 61.9 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-B-CELL;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC015765; AAH15765.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000832; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 13.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
 KW DNA-binding; Hypothetical protein; zinc-finger.
 SQ SEQUENCE 536 AA; 61932 MW; 3DC80FAE53F8CC00 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 536;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 50 PDLITCLE 57

RESULT 13
 ID 087IH2 PRELIMINARY; PRT; 132 AA.
 AC 087IH2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ATLESS RASGEF.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachbat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Lehmann B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC116548; AAL99317.1; -
 SQ SEQUENCE 132 AA; 14408 MW; 9AD15CE58D7666F8 CRC64;

Query Match 3.2%; Score 7; DB 5; Length 132;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 FSSATDQ 73
 |||||||

DB 57 FSSATDQ 63

RESULT 14
 09PPN3

ID Q9PPN3 PRELIMINARY; PRT: 135 AA.
 AC Q9PPN3;
 DT 01-OCT-2000 (TRMBLrel. 15, Created)
 DT 01-OCT-2000 (TRMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
 DE Putative ATP /GTP-binding protein.
 GN C10668.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 CC Campylobacter.
 OX NCBI_TaxID:197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCCTC 11168;
 RX MEDLINE-20150912; PubMed-10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
 RA Jørgensen K., Kariyasek A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrall B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL: AL139076; CAB72945.1;
 DR InterPro: IPR003442; UPP0079.
 DR Pfam: PF02367; UPP0079; 1.
 DR TIGRfams: TIGR00150; UPP0079; 1.
 KW Complete proteome.
 SQ SEQUENCE 135 AA; 15597 MW; 0E50BF8FDCD34E69 CRC64;

Query Match 3.28; Score 7; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 EGLANG 201
 |||||
 Db 80 EGLANG 86

RESULT 15

Q8TRK7 PRELIMINARY; PRT: 156 AA.
 ID Q8TRK7;
 AC Q8TRK7;
 DT 01-JUN-2002 (TRMBLrel. 21, Created)
 DT 01-JUN-2002 (TRMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
 DE Hypothetical protein MA3428.
 GN MA3428.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 CC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID:2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE-21929760; PubMed-11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels S., Smirnov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearlano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE011048; AM06795.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 156 AA; 17812 MW; 184878A5AF292852 CRC64;

Query Match 3.28; Score 7; DB 17; Length 156;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 80 SWAYPKE 86
 |||||
 Db 59 SWAYPKE 65

Search completed: May 15, 2003, 13:24:11
 Job time : 21.9089 secs

COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-77

Query Match 3.2%; Score 7; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 186 LOSKGIS 192
|||
DB 241 LOSKGIS 247

RESULT 3
US-09-175-828-77
Sequence 77, Application US/09175828
Patent No. 6221543
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175,828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-175-828-77

Query Match 3.2%; Score 7; DB 4; Length 370;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 186 LOSKGIS 192
|||
DB 241 LOSKGIS 247

RESULT 4
US-09-334-601-4
Sequence 4, Application US/09334601
Patent No. 6280989
GENERAL INFORMATION:
APPLICANT: Kapitonov, Dmitri
APPLICANT: YU, Robert
TITLE OF INVENTION: NOVEL STALYLTTRANSFERRASES
FILE REFERENCE: YCUIP-6
CURRENT APPLICATION NUMBER: US/09/334,601
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 94
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 4
LENGTH: 414
TYPE: PRT
ORGANISM: Murinae gen. sp.
US-09-334-601-4

Query Match 3.2%; Score 7; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 94 VYVLFKS 100
|||
DB 261 VYVLFKS 267

RESULT 5
US-09-186-276B-56
Sequence 56, Application US/09186276B
Patent No. 6388173
GENERAL INFORMATION:
APPLICANT: Bentley, Philip
APPLICANT: DiLaurenzio, Laura
APPLICANT: Wyszocka-Dillier, Joanna
APPLICANT: Malamy, Jocelyn E.
APPLICANT: Pysh, Leonard
APPLICANT: Helariutta, Yrjo
TITLE OF INVENTION: Sarcetow Gene, Promoter and Uses Thereof

FILE REFERENCE: 5914-075-999
CURRENT APPLICATION NUMBER: US/09/186,276B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(524)
OTHER INFORMATION: Xaa - Any Amino Acid
US-09-186-276B-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 6
US-08-842-445-56
Sequence 56, Application US/08842445A
Patent No. 6441270
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-056-999
CURRENT APPLICATION NUMBER: US/08/842,445A
CURRENT FILING DATE: 1997-04-24
EARLIER APPLICATION NUMBER: 08/638,617
EARLIER FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Plant
US-08-842-445-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 7
US-09-186-188B-56
Sequence 56, Application US/09186188B
Patent No. 6455672
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-074-999
CURRENT APPLICATION NUMBER: US/09/186,188B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26

NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(524)
OTHER INFORMATION: Xaa - Any Amino Acid
US-09-186-188B-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 8
US-08-781-891-75
Sequence 75, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-Bn
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620Leuburg, Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 607 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-75

Query Match 3.2%; Score 7; DB 3; Length 607;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LOSKGIS 192
DB 255 LOSKGIS 261

RESULT 9
US-08-796-414B-3
Sequence 3, Application US/08796414B
Patent No. 5876987
GENERAL INFORMATION:
APPLICANT: Wendy C. Champness, Paul Brian
APPLICANT: and Todd B. Anderson
TITLE OF INVENTION: METHOD, DNA AND BACTERIA
TITLE OF INVENTION: FOR HYPERPRODUCTION OF AN
TITLE OF INVENTION: ANTIBIOTIC DUE TO
TITLE OF INVENTION: DISRUPTION OF AN ADSA
NUMBER OF INVENTION: 7
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch,
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS 5.00
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/796, 414B
FILING DATE: February 6, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-297
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5876987e
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 103
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Amino Acid
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORGANISM: Streptomyces coelicolor
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: unicellular microorganisms
ORGANELLE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
FEATURE:
NAME/KEY: D7
LOCATION:
IDENTIFICATION METHOD: deduced
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:

ISSUE:
PAGES:
DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-796-414B-3

Query Match 2.7%; Score 6; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 155 TKDGYR 160
DB 89 TKDGYR 94

RESULT 10
US-08-241-853-13
Sequence 13, Application US/08241853
Patent No. 5693488
GENERAL INFORMATION:
APPLICANT: Fang, Kathy S.
APPLICANT: Hanafusa, Hidesaburo
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE
TITLE OF INVENTION: AND METHODS OF USE THEREOF
NUMBER OF INVENTION: 37
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,853
FILING DATE: 12-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-241-853-13
Query Match 2.7%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 35 EKITYV 40
DB 84 EKITYV 89

RESULT 11
US-08-850-917-13
Sequence 13, Application US/08850917
Patent No. 5854045

GENERAL INFORMATION:
APPLICANT: Fang, Kathy S.
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE
TITLE OF INVENTION: AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,917
FILING DATE: 02-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/241,853
FILING DATE: 12-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-850-917-13

Query Match
Best Local Similarity 2.7%; Score 6; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 EXITIV 40
|||||
DB 84 EXITIV 89

RESULT 12
US-08-887-352B-11
Sequence 11, Application US/08887352B
Patent No. 5994511
GENERAL INFORMATION:
APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
TITLE OF INVENTION: Improving Polypeptides
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,352B
FILING DATE: 03-Jul-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-887-352B-11

Query Match
Best Local Similarity 2.7%; Score 6; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
|||||
DB 98 RGSHPF 103

RESULT 13
US-08-887-352B-12
Sequence 12, Application US/08887352B
Patent No. 5994511
GENERAL INFORMATION:
APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
TITLE OF INVENTION: Improving Polypeptides
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,352B
FILING DATE: 03-Jul-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-887-352B-12

Query Match
Best Local Similarity 2.7%; Score 6; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
|||||

Db 98 RGSHPF 103

```
RESULT 14
US-09-109-207C-11
; Sequence 11, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 11
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-114
; OTHER INFORMATION: Heavy chain sequence derived from MAE11
US-09-109-207C-11
```

Query Match 2.7%; Score 6; DB 4; Length 114;
Best local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
Db 98 RGSHPF 103

```
RESULT 15
US-09-109-207C-12
; Sequence 12, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 12
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-114
; OTHER INFORMATION: Heavy chain sequence derived from MAE11
US-09-109-207C-12
```

Query Match 2.7%; Score 6; DB 4; Length 114;
Best local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
Db 98 RGSHPF 103

Search completed: May 15, 2003, 13:24:51
Job time : 14.0644 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 15, 2003, 13:22:46 ; Search time 11.6892 Seconds
(without alignments)
1807.034 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 BELGDGCHIVTSQDSGTMT.....NGVLSRHGSLSEKRLFTTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 362588 seqs, 96450795 residues

Word size : 0

Total number of hits satisfying chosen parameters: 204442

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	503	US-10-003-132-4	Sequence 4, Appl1
2	148	67.6	503	US-09-823-038A-51	Sequence 51, Appl1
3	100	45.7	421	US-10-003-132-6	Sequence 6, Appl1
4	28	12.8	458	US-09-759-130B-76	Sequence 76, Appl1
5	28	12.8	681	US-09-759-130B-75	Sequence 75, Appl1
6	7	3.2	359	US-10-309-389-2	Sequence 2, Appl1
7	7	3.2	370	US-09-753-143-77	Sequence 77, Appl1
8	7	3.2	385	US-09-712-363-148	Sequence 148, Appl1
9	7	3.2	524	US-10-253-007-56	Sequence 56, Appl1
10	7	3.2	604	US-10-045-170A-1	Sequence 1, Appl1
11	7	3.2	610	US-09-815-242-10414	Sequence 10414, A
12	6	2.7	100	US-09-864-761-47749	Sequence 47749, A
13	6	2.7	104	US-09-911-777-5	Sequence 5, Appl1
14	6	2.7	107	US-09-864-761-47562	Sequence 47562, A
15	6	2.7	108	US-09-910-150-21	Sequence 21, Appl1
16	6	2.7	110	US-09-864-761-42215	Sequence 42215, A
17	6	2.7	112	US-09-910-150-19	Sequence 19, Appl1
18	6	2.7	114	US-09-920-171-11	Sequence 11, Appl1
19	6	2.7	114	US-09-920-171-12	Sequence 12, Appl1

20	6	2.7	118	10	US-09-910-150-23	Sequence 23, Appl1
21	6	2.7	121	10	US-09-920-171-2	Sequence 2, Appl1
22	6	2.7	121	10	US-09-920-171-3	Sequence 3, Appl1
23	6	2.7	123	1	US-08-899-112-10	Sequence 10, Appl1
24	6	2.7	123	1	US-09-796-692-2374	Sequence 2374, Ap
25	6	2.7	123	9	US-10-040-862-2374	Sequence 2374, Ap
26	6	2.7	134	9	US-09-925-179-3	Sequence 3, Appl1
27	6	2.7	134	10	US-09-802-077-3	Sequence 3, Appl1
28	6	2.7	134	10	US-09-802-096-3	Sequence 3, Appl1
29	6	2.7	134	10	US-09-815-242-10513	Sequence 10513, A
30	6	2.7	148	9	US-09-738-626-6992	Sequence 6992, Ap
31	6	2.7	148	9	US-09-925-299-800	Sequence 800, App
32	6	2.7	148	10	US-09-925-299-800	Sequence 800, App
33	6	2.7	150	10	US-09-877-156-25	Sequence 25, Appl1
34	6	2.7	150	12	US-10-116-378-28	Sequence 28, Appl1
35	6	2.7	152	9	US-09-779-050X-24	Sequence 24, Appl1
36	6	2.7	157	9	US-09-803-327A-7	Sequence 7, Appl1
37	6	2.7	157	9	US-10-043-432-1	Sequence 1, Appl1
38	6	2.7	157	10	US-09-756-301A-1	Sequence 1, Appl1
39	6	2.7	157	10	US-09-927-703-1	Sequence 1, Appl1
40	6	2.7	157	10	US-09-854-280-19	Sequence 19, Appl1
41	6	2.7	157	10	US-09-934-465-13	Sequence 13, Appl1
42	6	2.7	157	10	US-09-766-535A-1	Sequence 1, Appl1
43	6	2.7	157	10	US-09-854-208-19	Sequence 19, Appl1
44	6	2.7	157	10	US-09-756-161A-1	Sequence 1, Appl1
45	6	2.7	157	12	US-10-010-229-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-10-003-132-4
Sequence 4, Application US/10003132
Publication No. US20020192750A1
GENERAL INFORMATION:
APPLICANT: Fox, Brian A.
APPLICANT: Gao, Zeren
APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCDB5
FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003,132
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/249,004
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 503
TYPE: PRT
ORGANISM: Mus musculus
US-10-003-132-4

Query Match	100.0%	Score 219;	DB 9;	Length 503;
Best Local Similarity	100.0%	Pred. No. 1.3e-204;		
Matches 219;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps 0;				
QY	1	EEI GDGCHIVTSQDSGTMTSKNYPGTYVTCETITVPKRRRLRLRGDINTSKTC	60	
DB	26	EEI GDGCHIVTSQDSGTMTSKNYPGTYVTCETITVPKRRRLRLRGDINTSKTC	85	
QY	61	ASDYLLFSATPDQGYPCGSMAWPKETRLNSNVTYLFKSGSHISGRGFLITAASSDHPD	120	
DB	86	ASDYLLFSATPDQGYPCGSMAWPKETRLNSNVTYLFKSGSHISGRGFLITAASSDHPD	145	
QY	121	LITCLERGSYFEFEKYSKFCPCACRDIAGDISGNFTKGYRDTSLCKAALHAGITDEIG	180	
DB	146	LITCLERGSYFEFEKYSKFCPCACRDIAGDISGNFTKGYRDTSLCKAALHAGITDEIG	205	
QY	181	GHTNLQSKGISHYEGILANGVLSRHGSLSEKRLFTTP	219	
DB	206	GHTNLQSKGISHYEGILANGVLSRHGSLSEKRLFTTP	244	

```
RESULT 2
US-09-823-038A-51
; Sequence 51, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Ormsted, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Muriel, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823.038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-51

Query Match          67.6%; Score 148; DB 10; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  EELDGGCHITSDSGMTSKNTPGTYPNVTCCKITTVKRGKLLIRLDNIESKTC 60
    |||||||
DB 26  EELDGGCHITSDSGMTSKNTPGTYPNVTCCKITTVKRGKLLIRLDNIESKTC 85
    |||||||

QY 61  ASDYLFFSSATDQGYPCGSAVPEKELRLNSNEVYLFKSGSHISGRFLTYASSDHPD 120
    |||||||
DB 86  ASDYLFFSSATDQGYPCGSAVPEKELRLNSNEVYLFKSGSHISGRFLTYASSDHPD 145
    |||||||

QY 121 LITCLERGSHYFEKYSKFCPCAGCDIA 148
    |||||||
DB 146 LITCLERGSHYFEKYSKFCPCAGCDIA 173
    |||||||

RESULT 3
US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003.132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Query Match          45.7%; Score 100; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.2e-89;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 DLITCLERGSHYFEKYSKFCPCAGCDIAGDISGNTKDGVRTSLCKAAIHAGITDEL 179
    |||||||
DB 100 DLITCLERGSHYFEKYSKFCPCAGCDIAGDISGNTKDGVRTSLCKAAIHAGITDEL 159
    |||||||

QY 180 GGHINLQSKGISHYEGILANGVLSRHGSLSEKRFLLFTTP 219
    |||||||
```

```
DB 160 GGHINLQSKGISHYEGILANGVLSRHGSLSEKRFLLFTTP 199
    |||||||

RESULT 4
US-09-759-130B-76
; Sequence 76, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirt, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NIM
; CURRENT APPLICATION NUMBER: US/09/759.130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-76

Query Match          12.8%; Score 28; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 4.5e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 GSHISGRGFLTYASSDHPDLITCLER 127
    |||||||
DB 100 GSHISGRGFLTYASSDHPDLITCLER 127
    |||||||

RESULT 5
US-09-759-130B-75
; Sequence 75, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
```

```

; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirtel, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350ANIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-759-130B-75

Query Match      12.8%; Score 28; DB 9; Length 681;
Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 GSGHISGRGFLTYASSDHPDLITCER 127
DB 100 GSGHISGRGFLTYASSDHPDLITCER 127

RESULT 6
US-10-309-389-2
; Sequence 2, Application US/10309389
; Publication No. US20030087396A1
; GENERAL INFORMATION:
; APPLICANT: Saito, Masaki
; TITLE OF INVENTION: Sialyltransferase and DNA encoding the same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/309,389
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: 09/425,488
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-184184
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/112,563
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: JP 11-148603
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 359
; TYPE: PRT

```

```

; ORGANISM: Mus musculus
; US-10-309-389-2

Query Match      3.2%; Score 7; DB 10; Length 359;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 VTVLFKS 100
DB 206 VTVLFKS 212

RESULT 7
US-09-753-143-77
; Sequence 77, Application US/09753143
; Patent No. US20020102550A1
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELIS, JAMES GERMAN, AND JOANNA
; GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/753,143
; FILING DATE: 02-Jan-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/175,828
; FILING DATE: 1998-10-20
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: <UNKNOWN>
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
; US-09-753-143-77

Query Match      3.2%; Score 7; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LOSKGIS 192
DB 241 LOSKGIS 247

```

```
RESULT 8
US-09-712-363-148
; Sequence 148, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-148
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 385;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 46 LRLRGD 52
|||
Db 280 LRLRGD 286
```

```
RESULT 9
US-10-253-007-56
; Sequence 56, Application US/10253007
; Publication No. US20030088073A1
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-074-999
; CURRENT APPLICATION NUMBER: US/10/253,007
; PRIOR FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: US/09/186,188
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
```

```
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-10-253-007-56
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 524;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 48 LRLGDLN 54
|||
Db 335 LRLGDLN 341
```

```
RESULT 10
US-10-045-170A-1
; Sequence 1, Application US/10045170A
; Patent No. US20020164809A1
; GENERAL INFORMATION:
; APPLICANT: Qiu, Yongchang
; APPLICANT: Wang, Jack
; APPLICANT: Hewick, Rodney
; TITLE OF INVENTION: ACID-LABILE ISOTOPE-CODED EXTRACTANT (ALICE) AND ITS USE IN GC
; FILE REFERENCE: G15412AUSA
; CURRENT APPLICATION NUMBER: US/10/045,170A
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242643
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Bovine Serum Albumin
US-10-045-170A-1
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 604;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 65 LLFSST 71
|||
Db 11 LLFSST 17
```

```
RESULT 11
US-09-815-242-10414
; Sequence 10414, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 10414
;; LENGTH: 610
;; TYPE: PRT
;; ORGANISM: Escherichia coli
US-09-815-242-10414

Query Match
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LQSKGIS 192
DB 258 LQSKGIS 264
```

```
RESULT 12
US-09-864-761-47749
; Sequence 47749, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
```

```
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 47749
;; LENGTH: 100
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL109809.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 0.88
;; OTHER INFORMATION: EST HUMAN HIT: AM406955.1, EVALUATE 1.00e-01
;; OTHER INFORMATION: SWISSPROT HIT: P01871, EVALUATE 7.00e-03
US-09-864-761-47749
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 LTVASS 116
DB 11 LTVASS 16
```

```
RESULT 13
US-09-911-777-5
; Sequence 5, Application US/09911777
; Patent No. US20020037852A1
; GENERAL INFORMATION:
; APPLICANT: BIOGEN, INC.
; APPLICANT: APOTECHE S.A.
; APPLICANT: BROWNING, Jeffrey
; APPLICANT: AMBROSE, Christine
; APPLICANT: MACKAY, Fabienne
; APPLICANT: TSCHOPE, Jurg
; APPLICANT: SCHNEIDER, Pascal
; TITLE OF INVENTION: BAFF, Inhibitors Thereof and Their Use
; FILE REFERENCE: A070 US
; CURRENT APPLICATION NUMBER: US/09/911,777
; CURRENT FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/117,169
; PRIOR FILING DATE: 1999-01-25
; PRIOR APPLICATION NUMBER: 60/143,228
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: PCT/US00/01788
; PRIOR FILING DATE: 2000-01-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-911-777-5

Query Match
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 LLANGV 202
DB 28 LLANGV 33

RESULT 14
US-09-864-761-47562
; Sequence 47562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
```

FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 47562
LENGTH: 107
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006023.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.6
OTHER INFORMATION: EST_HUMAN HIT: AW799294.1, EVALUATE 2.00e-47
OTHER INFORMATION: SWISSPROT HIT: P23201, EVALUATE 8.80e-01
US-09-864-761-47562

Query Match 2.7%, Score 6, DB 10, Length 107;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 184 NLSOK 189
Db 18 NLSOK 23

RESULT 15
US-09-910-150-21
Sequence 21, Application US/09910150
Patent No. US2002008698A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
APPLICANT: Rudolph-Owen, Laura
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Tsai, Fong Ying

TITLE OF INVENTION: 13237, 18480, 2245 OR 16228 NOVEL HUMAN
FILE REFERENCE: 38155-20020.00
CURRENT APPLICATION NUMBER: US/09/910,150
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/219,028
PRIOR FILING DATE: 2000-07-18
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 108
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid
US-09-910-150-21

Query Match 2.7%, Score 6, DB 10, Length 108;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 198 LANGVL 203
Db 100 LANGVL 105

Search completed: May 15, 2003, 13:26:36
Job time : 12.6892 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 45.4269 Seconds
(without alignments)
1475.448 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPGVTLALFVCAAP.....YSAPRNGIAPLNOTAMTALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2659	99.6	503	21	AA19126
2	1924.5	72.1	715	22	AAU00670
3	1309.5	49.0	539	22	AAU00630
4	1309.5	49.0	586	22	AAU00629
5	1154.5	43.2	487	22	AAU00628
6	715	26.8	398	23	AAE22715
7	715	26.8	398	23	AAU79459
8	553.5	20.7	583	22	AAE75450
9	532	19.9	385	23	AAE22716
10	532	19.9	385	23	AAU79460

11	532	19.9	385	23	ABB97386
12	523	19.6	365	23	AAE22721
13	436.5	16.3	669	21	AAV70539
14	181.5	6.8	671	23	AAE14565
15	181.5	6.8	686	23	AAE14568
16	177.5	6.6	686	23	AAE14564
17	170.5	6.4	75	22	ABB42217
18	170.5	6.4	75	22	AAW63102
19	170.5	6.4	75	22	AAW75913
20	170.5	6.4	75	22	AAW36024
21	170.5	6.4	75	23	ABG45349
22	170.5	6.4	415	20	AAI13361
23	170.5	6.4	415	21	AAW24404
24	170.5	6.4	415	22	AAW88559
25	169	6.3	372	22	AAW80229
26	168	6.3	415	22	AAW88415
27	164.5	6.2	415	20	AAV31743
28	164.5	6.2	419	22	ABG22064
29	162.5	6.1	1733	23	ABW6023
30	162.5	6.1	3631	22	ABG28277
31	161.5	6.0	3623	20	AAV27020
32	159.5	6.0	110	20	AAV27042
33	157	5.9	315	22	AAW75323
34	154	5.8	464	22	AAW93631
35	153.5	5.7	728	22	AAW47559
36	153.5	5.7	689	22	AAW85060
37	150.5	5.6	690	22	AAU00713
38	150.5	5.6	690	22	AAU00717
39	150.5	5.6	717	22	AAU00196
40	150.5	5.6	718	22	AAU00200
41	150.5	5.6	1083	22	AAU00714
42	150.5	5.6	1084	22	AAU00718
43	150.5	5.6	1111	22	AAU00197
44	150.5	5.6	1112	22	AAU00201
45	150.5	5.6	1192	22	AAU00712

ALIGNMENTS

RESULT 1

ID	AA19126	standard; Protein; 503 AA.
XX	AA19126	
AC	AA19126	
XX		
DT	19-FEB-2001	(first entry)
XX		
DE	Polypeptide isolated from lymph node stromal cells of fsn -/- mice.	
KW	Lymph node stromal cell; fsn -/- mice; inflammatory disorder;	
KW	immune system disorder; cancer; viral infection; HIV infection;	
KW	blood vessel growth; tumour necrosis factor disorder; arthritis;	
KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder;	
KW	cardiac failure.	
XX		
OS	Mus sp.	
XX		
PN	W0200058463-A1.	
PD	05-OCT-2000.	
XX		
PF	18-FEB-2000; 2000MO-NZ00015.	
XX		
PR	25-MAR-1999; 99US-0276268.	
PR	26-AUG-1999; 99US-0383586.	
XX		
PA	(GENE-) GENESIS RES & DEV CORP LTD.	
XX		
PI	Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;	
PI	Murison JG;	
XX		
DR	WPI; 2000-664924/64.	

DR N-PSDB: AAA96736.

XX polypeptide expressed in mammalian fan -/- lymph node stromal cells,
 PT useful for modulating growth of blood cells, for treating inflammatory
 PT and tumour necrosis factor-mediated disorders, cancer and viral
 PT disorders

PS Claim 1; Page 68-69; 75pp; English.

XX The present sequence represents a polypeptide sequence which is
 CC isolated from lymph node stromal cells of fan -/- mice. The
 CC polynucleotides and their polypeptides are useful for treating an
 CC inflammatory disorder, disorder of immune system and cancer selected
 CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
 CC viral disorder, in particular HIV infection and for modulating the
 CC growth of blood vessels. The polypeptides are useful for treating a
 CC tumour necrosis factor (TNF) mediated disorder, such as those selected
 CC from arthritis, inflammatory bowel disease and cardiac failure and a
 CC fibroblast growth factor-mediated disorder. It is also useful in assays
 CC to determine biological activity, to raise antibodies, to isolate
 CC corresponding ligands or receptors, to quantify levels of protein or
 CC cognate corresponding ligand or receptors, as anti-inflammatory agents,
 CC and in compositions for the treatment of skin, connective tissue and
 CC immune system diseases. The polynucleotide is useful as marker for
 CC tissue, as a chromosome marker or tags in the identification of a
 CC genetic disorder.

XX Sequence 503 AA:

Query Match 99.6%; Score 2659; DB 21; Length 503;

Best Local Similarity 99.6%; Pred. No. 2,6e-243;

Matches 501; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGTAGGSPVATLAFVACAPRLQAEEELGDCGHIYTSQDSGTMTSKNPGTYPYTACE 60
 DB 1 MGTAGGSPVATLAFVACAPRLQAEEELGDCGHIYTSQDSGTMTSKNPGTYPYTACE 60
 QY 61 KTTTPKRRRLRLRGDNIETSKASDYLLFSSATDQYGPYCGSWAYPKELRLNSNFT 120
 DB 61 KTTTPKRRRLRLRGDNIETSKASDYLLFSSATDQYGPYCGSWAYPKELRLNSNFT 120
 QY 61 KTTTPKRRRLRLRGDNIETSKASDYLLFSSATDQYGPYCGSWAYPKELRLNSNFT 120
 DB 61 KTTTPKRRRLRLRGDNIETSKASDYLLFSSATDQYGPYCGSWAYPKELRLNSNFT 120
 QY 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACRODIAGDISGNT 180
 DB 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACRODIAGDISGNT 180
 QY 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACRODIAGDISGNT 180
 DB 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACRODIAGDISGNT 180
 QY 181 KDGYRDTSLCKAAIHAGITDELGHINLQSKGISHYEGGLANGVLSRHGSLSEKRL 240
 DB 181 KDGYRDTSLCKAAIHAGITDELGHINLQSKGISHYEGGLANGVLSRHGSLSEKRL 240
 QY 241 FTTPCMNTTVAIPSVIFAILLTGMSGFATCRKRRKGNPNVSDAOKTGCKWKIKYF 300
 DB 241 FTTPCMNTTVAIPSVIFAILLTGMSGFATCRKRRKGNPNVSDAOKTGCKWKIKYF 300
 QY 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTTEY 360
 DB 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTTEY 360
 QY 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTTEY 360
 DB 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTTEY 360
 QY 361 RYNTFASGHYDCPHRPHRHEVALPTTHSEPEYATPIVENHLLRAHTFTSQSGRYVGGPP 420
 DB 361 RYNTFASGHYDCPHRPHRHEVALPTTHSEPEYATPIVENHLLRAHTFTSQSGRYVGGPP 420
 QY 421 TRKHSHSSGGPPRATGANGVESYQRPAPKPYGGGYDPAASSFLSDRPAQSQMSTSG 480
 DB 421 TRKHSHSSGGPPRATGANGVESYQRPAPKPYGGGYDPAASSFLSDRPAQSQMSTSG 480
 QY 481 DDGYAPRNGLAFLNQTAMTALL 503
 DB 481 DDGYAPRNGLAFLNQTAMTALL 503

RESULT 2
 AAU00670
 ID AAU00670 standard; Protein: 715 AA.

XX AAU00670;
 XX 07-SEP-2001 (first entry)
 DE Human TANGO 229 polypeptide.

XX Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrate; blood; serum.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..34

FT /note= "Signal peptide"

FT Domain 35..455

FT /note= "Extracellular domain"

FT Protein 35..715

FT /note= "Mature human TANGO 229"

FT Domain 456..480

FT /note= "Transmembrane domain"

FT Domain 481..715

FT /note= "Cytoplasmic domain"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

PF 19-OCT-1999; 99US-0420707.

PR (MILL-) MILLENNIUM PHARM INC.

PA Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

PI WPI: 2001-308477/32.

XX N-PSDB: AAS00660.

DR New isolated nucleic acid molecule for diagnosis, prevention, and

XX therapy of human and other animal disorder, or as modulating agent for

PT regulating cellular processes

PS Claim 8; Fig 1; 26pp; English.

XX The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration
 CC of e.g. T cells and cells of the heart, liver, pancreas, placenta,
 CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
 CC blood leukocyte, bone marrow or thymus tissue. They can be used as
 CC modulating agents for regulating cellular processes, thus, the proteins
 CC and their associated nucleic acids can be used to prognosticate, prevent,
 CC diagnose, or treat disorders associated with physiological processes.
 CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
 CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
 CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
 CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
 CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
 CC muscular dystrophy. Antibodies to disorders such as these can be made by
 CC providing a polypeptide of the invention to an immuno-competent
 CC vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA:

Query Match 72.1%; Score 1924.5; DB 22; Length 715;
 Best Local Similarity 55.8%; Pred. No. 2.4e-173;

CC Inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These
 CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.

XX Sequence 487 AA;

Query Match 43.2%; Score 1154.5; DB 22; Length 487;
 Best Local Similarity 48.8%; Pred. No. 1.7e-100;
 Matches 237; Conservative 24; Mismatches 26; Indels 199; Gaps 1;

QY 44 MTSKNPPTGTPNTVCEKITTPVCKGRLLRLDLDNTESTCASDYLTFSSATQVPPYC 103
 DB 1 MTSKNPPTGTPNTVCEKITTPVCKGRLLRLDLDNTESTCASDYLTFSSATQVPPYC 60
 QY 104 GSNAPVPELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLITCLERSHYPEEKYSK 163
 DB 61 GSNAPVPELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLITCLERSHYPEEKYSK 120
 QY 164 FCPAGCRDIAGDISGNTKDGYPDTSLCKAAIHAGITDELGCHINLQSGKISHYGLL 223
 DB 121 FCPAGCRDIAGDISGNTKDGYPDTSLCKAAIHAGITDELGCHINLQSGKISHYGLL 180
 QY 224 ANGVLSHSGLSSEKRRPF----- 241
 DB 181 ANGVLSHSGLSSEKRRPF----- 240
 QY 242 ----- 241
 DB 241 QDGPSSWASDSSNNHKKRMLLEIDGKKKIRGTGSGSNFNFYVSVNFKNN 300
 QY 242 ----- 241
 DB 301 SKWKTYGIVNNEKVFQGSNFRDPYQNNFIPYAVYRVVPQWHORIALKEVLIGC 360
 QY 242 -----TPGMNTTVAIPSVFIALLT 264
 DB 361 QITQGNDSLWKRKTSOSTSVTKKEDFTTRIPDSEETSTGINTTVAIPVLAIVLVAFA 420
 QY 265 GMGIFALICRRKKRGPNVSADQKTCMKQIKXPFAHOSHTEFTISYDNEKEWTKRLD 324
 DB 421 GMGIFALICRRKKRGPNVSADQKTCMKQIKXPFAHOSHTEFTISYDNEKEWTKRLD 480
 QY 325 ITSDMA 330
 DB 481 ITSDMA 486

RESULT 6
 AAE22715
 ID AAE22715 standard; Protein; 398 AA.
 XX
 AC AAE22715;
 XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy1 protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW opthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;

KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virocid; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 OS Homo sapiens.
 XX
 PN MO200222815-A1.
 XX
 PD 21-MAR-2002.
 XX
 PE 12-SEP-2001; 2001MO-US28488.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 XX
 PR 06-SEP-2001; 2001US-317902P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT;
 XX
 DR WPI: 2002-393966/42.
 XX
 DR N-PSDB; AAD35992.
 XX
 PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 PS Claim 3; Page 123-125; 152pp; English.

CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy1
 CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.
 XX

SO Sequence 398 AA;

Query Match 26.8%; Score 715; DB 23; Length 398;
 Best Local Similarity 54.3%; Pred. No. 6.7e-59;
 Matches 159; Conservative 36; Mismatches 60; Indels 38; Gaps 8;

QY 29 GDGCGHIVTSODSGTMTSKNYPGTPYPTVCEKITTPVCKGRLLRLDLDNTESTCASD 88
 DB 4 GDGCGHIVTSODSGTMTSKNYPGTPYPTVCEKITTPVCKGRLLRLDLDNTESTCASD 63
 QY 89 YLFFSSATDQYGPYCGSNAPVPELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLIT 148
 DB 64 YLFFSSATDQYGPYCGSNAPVPELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLIT 121
 QY 149 CLENGSHYFEKYSK-FCPAGCRDIAGDISGNTKDGYPDTSLCKAAIHAGITDELGCH 207

Db 122 QGDRSEKTLDOQSRTFLATGTFVFKDSFS---TDG---TSLCKRAAHAGIADDELGQ 175
QY 208 INLQSKGISHEGLANGVLSRHGSLSEKRF---LFTPGMNTTYAIPSVIRIALL 263
Db 176 ISVLQKRGISREGLANGVLSREFEIRFQELFSSVLYSMGNTYHAV-----TELMF 228
QY 264 TGMGIFAIORKR-----KKGNPYVSADAK-----TCCKMOKY 298
Db 229 PHMIVHSGKTRREGSIAAESEGVPLVLYIQKQELVODLVIVATVGCGRSLSF 281
RESULT 7
AAU79459
ID AAU79459 standard; Protein; 398 AA.
AC AAU79459;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human Neuropilin-Hy1.
XX
KW Human: neuropilin-Hy1; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.
XX
OS Homo sapiens.
XX
PN WO200222780-A2.
XX
PD 21-MAR-2002.
XX
PF 11-SEP-2001; 2001WO-US28590.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-0659671.
XX
PA (TANG/) TANG T Y.
XX
PI Tang TY;
XX
DR WPI: 2002-351881/38.
DR N-PSDB; ABK49565.
XX
PT New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation
XX
PS Claim 3; Page 118-120; 144pp; English.
XX
OS The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for

CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy1.
XX
SQ Sequence 398 AA;
XX
Query Match 26.8%; Score 715; DB 23; Length 398;
Best Local Similarity 54.3%; Pred. No. 6.7e-59;
Matches 159; Conservative 36; Mismatches 60; Indels 38; Gaps 8;
QY 29 GDGCHIVTSODSGTMTSKNTPGTYPNYTCERKITVYKGRLLIRLGDNLNIESKTCASD 88
Db 4 GDGCHIVTYODSGTMTSKNTPGTYPNYTCERKITVYKGRLLIRLGDNLNIESKTCASD 63
QY 89 YLFFSARDQYGYCGSMAPVELNLNSNEVTLFKSGSHISGRGFLTYASSDHPDLIT 148
Db 64 YLFTSSSDQYGYCGSMVTPRELNLNTEVTVRESGSHISGRGFLTYASSDHPD--S 121
QY 149 CLERSHVFEEYK-RPCPACGRDIAGDISGTRKGYRDTSLCKRAAHAGITDELGCH 207
Db 122 QGDRSEKTLDOQSRTFLATGTFVFKDSFS---TDG---TSLCKRAAHAGIADDELGQ 175
QY 208 INLQSKGISHEGLANGVLSRHGSLSEKRF---LFTPGMNTTYAIPSVIRIALL 263
Db 176 ISVLQKRGISREGLANGVLSREFEIRFQELFSSVLYSMGNTYHAV-----TELMF 228
QY 264 TGMGIFAIORKR-----KKGNPYVSADAK-----TCCKMOKY 298
Db 229 PHMIVHSGKTRREGSIAAESEGVPLVLYIQKQELVODLVIVATVGCGRSLSF 281
RESULT 8
AAG75450
ID AAG75450 standard; Protein; 583 AA.
AC AAG75450;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6214.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCL INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-23537/24.
DR N-PSDB; AAH34855.
XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11, Page 7657-7660; 9803pp; English.
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence 583 AA;
 SQ
 Query Match 20.7%; Score 553.5; DB 22; Length 583;
 Best Local Similarity 30.3%; Pred. No. 2.6e-43;
 Matches 163; Conservative 82; Mismatches 194; Indels 99; Gaps 19;
 XX
 OY 8 PSVALTFAVCAPLRLQAEELGDCGHIYVSODSGTMSKNPQYPTVYCEKIIYTPK 67
 DB 26 PLFLLTLLVLLLEDDAGAGGDCGHTVGLPESGTLTISINYPOTYNSFCWEIIRYKM 85
 OY 68 GKRLILRLGLNIE-SKTQASDYILPSS-----ATDYGPICG-SWAVPKRLNLSNEYTV 121
 DB 86 GERVIRKFGDIEDSDSCHFNRYIRYNGIVSFTKIGKTCGLGLQNMHSIESKGNITL 145
 OY 122 LFKSGSHISGRGFLTYAASDHPDLITCLERGSHYFEKYSKFCPCACRIADISGNTK 181
 DB 146 LFMGSHVSGRGLASVLDKODLITCLDTPASHFLPERSKYCPACCLPFAFISGTP 205
 OY 182 DGYRDTSLCKAAIHAGITDELQGINLQSKGISHYEGLANGVLRSRHSLEKRFLE 241
 DB 206 HGRDSSPLCMAGVHAGVSNLTGQISVYISKIPYESSLANNVTSVGHLSLSTFTF 265
 OY 242 TTP-----GMITTVAIPIVPIALL-----LTGM-GTFAICRRKKKGNPYASADAKT 290
 DB 266 KTGCGYCTLMEGAVIADPQITASSVLEWTDHTGOENSMKPKARLRKPPPMALAFATDE 325
 OY 291 GCMKQIKYPPARHQSFEFTSYDNEKEMTKLIDITSDMADYQPLMIGTIV-----AR 345
 DB 326 YQWLQ-----IDLNKKKIT--GITTTGITWENNYVSAVRILYSDDGQ 368
 OY 346 KGSFTFMDTEFEVRNTEASGHYDCPHRPGRHRYALPLTHSEPEATFIVERHLLRAH 405
 DB 369 KMVYVRPGEVQDKI-----FQGNKDY-HQDVRNKF-LP-----PLIAR-FIRVN 410
 OY 406 TSTGSGRY-----PCRPRIKSHSHSGGPPATGATQVESYO 444
 DB 411 PTWQOKIAKMKELGCOPIPKGRPPRLTPRPPRNSNDKNTTA--PPRIAKRAKRF 468
 OY 445 RPASPKRVGGYDUPAASSFLDSDPASOSQMTSGGDDGYSAAPNGLAGLNOTAMTAL 502
 DB 469 QPIQPR-----SSNEF-----PAOTEGTTASPD-----IRNTVTPTNVTQVAL 507

XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy2 protein.
 KW Human: neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; prozoocidic;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnerable; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Signal_peptide
 FT 61..385
 FT Protein /note="Human mature neuropilin-Hy2 protein"
 FT
 XX NO200222815-A1.
 XX
 XX 21-MAR-2002.
 XX
 XX 12-SEP-2001; 2001MO-US28488.
 XX
 XX 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-317902P.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT;
 XX
 XX WPI, 2002-393966/42.
 DR N-PDB; AAD35994.
 XX
 PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 PS Claim 3; Page 128-130; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes

CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.
XX
XX
SQ Sequence 385 AA;

Query Match 19.9%; Score 532; DB 23; Length 385;
Best Local Similarity 45.5%; Pred. No. 1.5e-41;
Matches 111; Conservative 38; Mismatches 89; Indels 6; Gaps 3;

OY 8 PSVLLAFVACAPLRLQAEELDGGCHVTSDSGTMTSKNPGYPPNTVCEKTIITPK 67
DB 2 PFLVLLLVLLLLLEDAGAQGGCGCHTVLPESGTLISINPOTYPNSTVCEWEIRYKM 61
OY 68 GKRLIRLGLDLINE-SKTCASDYLLFSS---ATDQGPYCG-SVAWPKELRLNSNEVTY 121
DB 62 GERVAKIKGDPDIEDSDSCHFNRYLRYNGICVSRFEIKGYCGIGLGMNHSISKNETTL 121
OY 122 LFKSGSHISGRGFLTYVSSDHPDLITCLERSHYFEERYKFCPCAGCDIAGDISGNTK 181
DB 122 LFMSSIHVSGRGFLASVYIDKQDITCLDTPASNLFEPFSKYCPAGCLLPAEISGITP 181
OY 182 DGYRDTSLCKRAIHAGIITDELGGHINLLOSKGISHYEGLLANGVLSRHGSLSEKRLF 241
DB 182 HGYRDSPLCMAGVAGVAVSNTLGGQISVVISKGIPIYESSIANNTSVVGHLSLSTP 241
OY 242 TTPG 245
DB 242 KTSG 245

RESULT 10
AAU79460
ID AAU79460 standard; Protein: 385 AA.

AAU79460;

15-JUL-2002 (first entry)

Human Neuropilin-Hy2.

Human; neuropilin-Hy2; chromosome 6q21; neuronal growth;
neurodegeneration; neurodegenerative disease; learning disorder;
memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
organ growth; nervous system lesion; cancer; cell proliferation;
cell differentiation; stem cell growth factor activity;
Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
reperfusion; food supplement; DNA microarray.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..20

Protein /label= Signal_peptide
/label= Mature_neuropilin_Hy2

MO200222780-A2.

21-MAR-2002.

11-SEP-2001; 2001WO-US285950.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-0659671.

(TANG/) TANG T Y.
Tang TY;

DR WPI: 2002-351881/38.
DR N-PSDB; ABR49567.
XX
XX
XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation.

Claim 3; Page 123-125; 144pp; English.

The invention relates to an isolated neuropilin-like polypeptide
including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
encoding the proteins and the coding regions of the cDNAs. Also included
is a nucleic acid array comprising the cDNAs attached to a surface used
for detecting full-matches or mismatches to the cDNAs. The genes
for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
The nucleic acid array is useful for detecting full-matches or mismatches
to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
in modulating neuronal growth, regenerative capacity, treating
neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy2.
XX

Sequence 385 AA;

Query Match 19.9%; Score 532; DB 23; Length 385;

Best Local Similarity 45.5%; Pred. No. 1.5e-41;

Matches 111; Conservative 38; Mismatches 89; Indels 6; Gaps 3;

OY 8 PSVLLAFVACAPLRLQAEELDGGCHVTSDSGTMTSKNPGYPPNTVCEKTIITPK 67
DB 2 PFLVLLLVLLLLLEDAGAQGGCGCHTVLPESGTLISINPOTYPNSTVCEWEIRYKM 61
OY 68 GKRLIRLGLDLINE-SKTCASDYLLFSS---ATDQGPYCG-SVAWPKELRLNSNEVTY 121
DB 62 GERVAKIKGDPDIEDSDSCHFNRYLRYNGICVSRFEIKGYCGIGLGMNHSISKNETTL 121
OY 122 LFKSGSHISGRGFLTYVSSDHPDLITCLERSHYFEERYKFCPCAGCDIAGDISGNTK 181
DB 122 LFMSSIHVSGRGFLASVYIDKQDITCLDTPASNLFEPFSKYCPAGCLLPAEISGITP 181
OY 182 DGYRDTSLCKRAIHAGIITDELGGHINLLOSKGISHYEGLLANGVLSRHGSLSEKRLF 241
DB 182 HGYRDSPLCMAGVAGVAVSNTLGGQISVVISKGIPIYESSIANNTSVVGHLSLSTP 241
OY 242 TTPG 245
DB 242 KTSG 245

RESULT 11
ABB97386
ID ABB97386 standard; Protein: 385 AA.

[illegible]

AAE22721
ID AAE22721 standard; Protein; 365 AA.
XX
XX AAE22721;
AC
XX
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 mature protein.
XX
XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
XX neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
XX thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
XX ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
XX wound healing; tissue repair; Parkinson's disease; Huntington's disease;
XX amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
XX cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
XX human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
XX systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
XX autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
XX myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
XX neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
XX neuroprotective; tranquilliser; virucide; antibacterial; cytostatic;
XX immunosuppressive.
XX
XX Homo sapiens.
OS
XX
XX WO200222815-A1.
PN
XX
XX 21-MAR-2002.
PD
XX
XX 12-SEP-2001; 2001WO-US28488.
PF
XX
XX 11-SEP-2000; 2000US-0659671.
PR
XX 06-SEP-2001; 2001US-317902P.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT;
PI
XX
XX WPI; 2002-393966/42.
DR
XX
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
PS
PS
XX
XX
XX Disclosure; Page 131-132; 152pp; English.

FH Key Location/Qualifiers
 FT Region 13.671
 FT /note- "This region is specifically claimed in
 claim 45"
 XX
 XX WO200206460-A2.
 PN 24-JAN-2002.
 PD
 XX
 XX 13-JUL-2001; 2001WO-DK00499.
 PF
 XX 13-JUL-2000; 2000DK-0001089.
 PR 01-JUN-2001; 2001DK-0000870.
 XX
 XX (JENS/) JENSENIUS J C.
 PA (THIE/) THIEL S.
 XX
 XX Jansenius JC, Thiel S;
 PI
 XX WPI; 2002-179791/23.
 DR
 XX
 XX Use of a polypeptide comprising amino acid sequence derived from
 PT mannan-binding lectin associated serine protease-2 (MASP-2) for
 PT producing pharmaceutical composition, to treat bacterial, fungal, viral
 PT infections
 XX
 XX Claim 45; Page 73-76; 76pp; English.
 PS
 XX
 XX The invention relates to use of a polypeptide derived from
 CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
 CC producing a pharmaceutical composition. MASP-2 is a complement-
 CC fixing enzyme and involved in lectin pathway of complement activation.
 CC The pharmaceutical composition comprising MASP-2 is useful for
 CC treating infections caused by microbes such as fungus, yeast,
 CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
 CC bacteria which are resistant to at least one antibiotic medication or
 CC multiresistant. The polynucleotide encoding MASP-2 is useful for
 CC treating patients deficient in MASP-2. The invention also discloses
 CC MASP-2 assays which are useful for determination of MASP-2 activity or
 CC levels in patients suffering from e.g. infections, inflammatory disorders
 CC and spontaneous recurrent abortion. The pharmaceutical composition
 CC comprising MASP-2 inhibitor is useful for treating inflammatory
 CC disorders. The present sequence is human mature MASP-2 protein.
 CC
 XX
 XX Sequence 671 AA;
 SQ
 Query Match 6.8%; Score 181.5; DB 23; Length 671;
 Best Local Similarity 24.4%; Pred. No. 6.9e-08;
 Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;
 OY 32 CGHIVTSDSGTWSKRNPGTYPCVCEKIIIVPKGRLLIL---RLGDLNIESKT-CAS 87
 DB 169 GSGOVFTORSGELSPREYPRPKLSCTVYSISLEGGFVITLDFVESFVETHPETLCPY 228
 OY 88 DYLFSSADDQYPPYCGSNAPVPEKRLNSNEVYLF---KSGSHISRGFLTYASDH- 143
 DB 229 DFLKIDTDRREHPCFGCK-TLPHRIETKSNVTITFTVTDSSGH---TGKRIHTSTAH 284
 OY 144 -----PDLITCLERSHFEKYSKFCPAGCRDIAGDISGNT-----KDGVRDLSL 190
 DB 285 CPTPAPMPKPNHVSPOAKYITLKDSFSIFCTGYELLQSHLPLKSFYAVCOKDGSMDRPM 344
 OY 191 CKRAIHAGIITDELGHINILQSGKISHYEGLL 223
 DB 345 ACSIVDCGPPDDLPSGRVEXYITGPVYTVYKAVI 377
 RESULT 15
 AAEL14568
 ID AAEL14568 standard; Protein; 686 AA.
 AC AAEL14568;
 XX

DT 17-MAY-2002 (first entry)
 XX
 DE Human MASP-2 protein, alternative version.
 XX
 XX Human: mannan-binding lectin associated serine protease-2; MASP-2; MBL;
 KW complement fixation; infection; microbe; retrovirus; HIV; abortion;
 KW human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..15
 FT /label= signal_peptide
 FT Protein 16..686
 FT /label= Mature_MASP-2_protein
 FT MISC-difference 155
 FT /note= "Encoded by CA"
 FT MISC-difference 156
 FT /note= "Encoded by C"
 FT MISC-difference 298
 FT /note= "Encoded by CAG"
 FT MISC-difference 299
 FT /note= "Encoded by CCG"
 XX
 PN WO200206460-A2.
 PD
 XX
 XX 24-JAN-2002.
 PD
 XX
 XX 13-JUL-2001; 2001WO-DK00499.
 PF
 XX 13-JUL-2000; 2000DK-0001089.
 PR 01-JUN-2001; 2001DK-0000870.
 XX
 XX (JENS/) JENSENIUS J C.
 PA (THIE/) THIEL S.
 XX
 XX Jansenius JC, Thiel S;
 PI
 XX WPI; 2002-179791/23.
 DR N-PSDB; AAD24224.
 XX
 XX Use of a polypeptide comprising amino acid sequence derived from
 PT mannan-binding lectin associated serine protease-2 (MASP-2) for
 PT producing pharmaceutical composition, to treat bacterial, fungal, viral
 PT infections
 XX
 XX Claim 41; Page 71-73; 76pp; English.
 PS
 XX The invention relates to use of a polypeptide derived from
 CC mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for
 CC producing a pharmaceutical composition. MASP-2 is a complement-
 CC fixing enzyme and involved in lectin pathway of complement activation.
 CC The pharmaceutical composition comprising MASP-2 is useful for
 CC treating infections caused by microbes such as fungus, yeast,
 CC retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic
 CC bacteria which are resistant to at least one antibiotic medication or
 CC multiresistant. The polynucleotide encoding MASP-2 is useful for
 CC treating patients deficient in MASP-2. The invention also discloses
 CC MASP-2 assays which are useful for determination of MASP-2 activity or
 CC levels in patients suffering from e.g. infections, inflammatory disorders
 CC and spontaneous recurrent abortion. The pharmaceutical composition
 CC comprising MASP-2 inhibitor is useful for treating inflammatory
 CC disorders. The present sequence is human MASP-2 protein.
 CC Note: The present sequence is stated as being the same as
 CC SEQ ID NO:2 shown in figure 6 of the specification (AAEL14564).
 CC However the sequences differ at various locations.
 CC
 XX
 XX Sequence 686 AA;
 SQ
 Query Match 6.8%; Score 181.5; DB 23; Length 686;
 Best Local Similarity 24.4%; Pred. No. 7.1e-08;
 Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;
 OY 32 CGHIVTSDSGTWSKRNPGTYPCVCEKIIIVPKGRLLIL---RLGDLNIESKT-CAS 87
 DB 169 GSGOVFTORSGELSPREYPRPKLSCTVYSISLEGGFVITLDFVESFVETHPETLCPY 228
 OY 88 DYLFSSADDQYPPYCGSNAPVPEKRLNSNEVYLF---KSGSHISRGFLTYASDH- 143
 DB 229 DFLKIDTDRREHPCFGCK-TLPHRIETKSNVTITFTVTDSSGH---TGKRIHTSTAH 284
 OY 144 -----PDLITCLERSHFEKYSKFCPAGCRDIAGDISGNT-----KDGVRDLSL 190
 DB 285 CPTPAPMPKPNHVSPOAKYITLKDSFSIFCTGYELLQSHLPLKSFYAVCOKDGSMDRPM 344
 OY 191 CKRAIHAGIITDELGHINILQSGKISHYEGLL 223
 DB 345 ACSIVDCGPPDDLPSGRVEXYITGPVYTVYKAVI 377
 RESULT 15
 AAEL14568
 ID AAEL14568 standard; Protein; 686 AA.
 AC AAEL14568;
 XX

```

OY 32 CGHITSQDSGTWMSKNTPGTYPNVTCEKIITVPKRLIL---RLGDLNIESKT-CAS 87
Db 184 CSGQVFTORSGEISPEYPRPYPKLSCTYSISLEGEFSVILDFVESFVETHPETLCBY 243
OY 88 DYLFSSATDOYGPYCGSMAYPKELRLINSNEVTLPF---KSGSHISGRGELLTYASDH- 143
Db 244 DFLKIOTDREHGPFGK-TLPHRIETKSNVTYITFVTDESGDH--TGWKIHYTSTAH 299
OY 144 -----PDLITCLERGRSHYEEKYSKFCPACGRDIAGDISGNT-----KGYRDTSL 190
Db 300 CPYMAPPENGHSPYQAKYILKDSFISFCETGYELLQGHLPKSFYAVCQKDGSMWRPMP 359
OY 191 CKAATHAGIITDELGCHINLQSKGISHEGLL 223
Db 360 ACSIVDCGPPDDLPSGRVEYITGPGVTYKAVI 392

```

Search completed: May 15, 2003, 13:20:46
 Job time : 49.4269 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 14.454 Seconds

(Without alignments)
3345.478 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671

Sequence: 1 MGTGAGGSPVTLALFAVCAP.....YSAPRNGLAPLQNTAMTALL 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : PIR_73:**

1: p1r1:**

2: p1r2:**

3: p1r3:**

4: p1r4:**

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	181.5	6.8	686	1 A59271	Ra-reactive factor
2	169.5	6.3	694	2 JG6554	complement subcomp
3	168	6.3	1524	2 T30337	polyprotein - Afri
4	167.5	6.3	3623	2 T08618	intrinsic factor-B
5	162.5	6.1	3623	2 T09456	intrinsic factor-B
6	158.5	5.9	695	1 S05008	complement subcomp
7	157	5.9	927	1 JQ0948	A5 antigen precurs
8	151.5	5.7	699	1 I54763	Ra-reactive factor
9	151	5.7	688	1 C1HRS	complement subcomp
10	150	5.6	1594	2 T30549	heparin - rabbit
11	147	5.5	705	1 C1HURB	complement subcomp
12	142.5	5.3	966	1 B58788	procollagen C-endo
13	142.5	5.3	991	2 I49540	procollagen C-endo
14	142	5.3	579	2 JG7629	membrane-type friz
15	142	5.3	2083	2 T42721	CRP-ductin-alpha p
16	141.5	5.3	1070	2 T31069	tollid-BMP-1 like
17	141	5.3	319	2 I51569	UVS-2 protein - Af
18	137	5.1	449	2 A55362	procollagen I C-pr
19	130.5	4.9	1290	2 A55190	ebnerin precursor
20	130.5	4.9	3871	2 T22812	hypothetical prote
21	127.5	4.8	707	2 JG2218	procollagen C-endo
22	127	4.8	855	2 JG7731	membrane-bound arg
23	127	4.8	855	2 JG7731	membrane-bound arg
24	126	4.7	402	2 JH0403	procollagen I C-pr
25	125.5	4.7	2403	2 A59386	sanko - human
26	125	4.7	1019	2 A38738	coagulation factor
27	124	4.6	1057	1 A39288	dorsal-ventral pat
28	123.5	4.6	823	1 A58788	procollagen C-endo
29	122	4.6	730	1 BMH01	procollagen C-endo

30	121.5	4.5	597	2 S71352	metalloproteinase
31	121.5	4.5	1004	2 T30338	oviductin (EC 3.4.
32	119	4.5	767	2 T30018	hypothetical prote
33	117.5	4.4	347	2 T20618	hypothetical prote
34	116	4.3	275	2 JG6506	tumor necrosis fac
35	115	4.3	277	2 A41735	hyaluronate-bindin
36	114	4.3	288	2 T33224	hypothetical prote
37	113.5	4.2	1064	2 A40136	fibropellin Ia - s
38	112.5	4.2	1464	2 S58984	development protei
39	111	4.2	846	2 H70599	hypothetical prote
40	111	4.2	1087	2 T31100	probable potassium
41	110.5	4.1	436	2 T23345	hypothetical prote
42	110	4.1	276	2 A47290	TSG-6 homolog PS4
43	110	4.1	504	2 S56745	mycin (clone PGM31
44	109.5	4.1	321	2 T33161	hypothetical prote
45	109.5	4.1	1571	2 T00062	hypothetical prote

ALIGNMENTS

RESULT 1

A59271 Ra-reactive factor (EC 3.4.21.-) 2 precursor - human

N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)

C:Species: Homo sapiens (man)

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000

C:Accession: A59271

R:Phlel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaebel, W.J.; Laursen, S.B.; Poulsen, Nature 386, 506-510, 1997

A:Title: A second serine protease associated with mannan-binding lectin that activate

A:Reference number: A59271; MUID:97242412; PMID:9087411

A:Accession: A59271

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-686 <JEN>

A:Cross-references: GB:Y09926; NID:94007626; PIDN:CAA71059.1; PID:94007627

A:Experimental source: tissue liver

A:Note: submitted to GenBank, December 1996

A:Note: parts of this sequence, including the amino end of the mature protein, were d

A:Gene: GDB:MASP2

A:Cross-references: GDB:6071500

A:Map position: 1p36.2-1p36.3

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto

C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serin

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-444,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>

F:142-180/Domain: C1r/C1s repeat homology <EGF>

F:184-293/Domain: C1r/C1s repeat homology <C1R2>

F:306-430/Domain: complement factor H repeat homology <FH1>

F:445-679/Domain: complement factor H repeat homology <FH2>

F:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,396-430,434-5

F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:444-445/Cleavage site: Arg-tile (autolytic) #status predicted

F:483,532,633/Active site: His, Asp, Ser #status predicted

Query Match 6.88; Score 181.5; DB 1; Length 686;

Best Local Similarity 24.44; Pred. No. 7.3e-06;

Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;

DB 32 CGHIVTSGDGTMTSKNVPGRVYVTCERKITTPKRRRLT---RGDNIESTK-CAS 87

DB 184 CCGGVFTORSGELSSPEPRPRPLSLSCYSISLEEGSVILDFVESFDVETHPETLCPY 243

QY 88 DYLFSSATDQYGPYCGSMWPKELRLNSNEVYLF--KSGSHISRGFLTYASSDH- 143

DB 244 DFLKIQDRREHGFCK-TLPHRIETKSNVTITFTVDESGDH---TGMKIHTSTAH 299

QY 144 -----PDLITCLERSHFFEEKYKFCPCAGCDINDIGISGNT-----KDGRTDSL 190

Db 300 CYPMPAPNGHVSPVQAKIILKSDSPSIFCETGVELLQGLPLKSTFAYCQKQKSDMRPMP 359
Qy 191 CKAALHAGITIDELGCHINLQSKGISHYEGIL 223
Db 360 ACSIVDCGPPDDLPSGRVRYITGPGVTKYKAVI 392

RESULT 2

JC6554
Complement subcomponent C 158AR.GIF (EC 3.4.21.42) precursor [similarity] - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 03-Jun-2002
C:Accession: JC6554
R:Sakai, H.; Nakashima, S.; Yoshimura, S.; Nishimura, Y.; Sakai, N.; Nozawa, Y.
Gene 209, 87-94, 1998

A:Title: Molecular cloning of a cDNA encoding a serine protease homologous to complement
A:Reference number: JC6554; MUID:98192519; PMID:9524231
A:Accession: JC6554
A:Molecule type: mRNA

A:Residues: 1-694 <SAK>

A:Cross-references: DDBJ:D88250; NID:93080541; PIDN:BAA25797.1; PID:93080542

C:Comment: This protein is involved in glial cell differentiation and cartilage remodeling

C:Genetics:

A:Gene: r-98P

C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H

C:Keywords: differentiation; glycoprotein; hydrolase; serine proteinase

F:1-31/Domain: signal sequence #status predicted <SIG>

F:12-694/Product: C1r/C1s repeat homology <C1R>

F:141-177/Domain: serine protease homolog #status predicted <MAT>

F:300-360/Domain: EGF homology <EGF>

F:444-681/Domain: complement factor H repeat homology <FHR>

F:180,412/Binding site: carboxyrate (Asn) (covalent) #status predicted

F:481,535,637/Active site: His, Asp, Ser #status predicted

Query Match 6.3%; Score 169.5; DB 2; Length 694;
Best Local Similarity 28.0%; Pred. No. 6.3e-05;

Matches 52; Conservative 31; Mismatches 78; Indels 25; Gaps 7;

Qy 15 FAVCAP---LRQAELGDCGHITVSQDSGMTSKNTPTGTPNTVCCKITTPKGRKL 71

Db 161 FCSCEPEYFLHDMRTCGNCGDVFALIGIASPNPNPENSRCYQIRLQEGFRL 220

Qy 72 IL--RLGDLNIE---SKTCASDYLFSATDQYGPCGS--WAVPEKRLNSNEVYLFKS 125

Db 221 VLTTRDFVDPADSEGCNDSLTPAAKNOQFPGCGFPGPLTKTOSMTLDIVFQT 280

Qy 126 GSHISRGFLTYASSDHPDLITC-----LENGSHYEEKYKFTCPAGCRDING 174

Db 281 DLTGKNGKWLKLY---HDDPIPCPEKISANSIWEPEKAKYFKDVVKITVDGEFVEVG 336

Qy 175 DISGNT 180

Db 337 NV-GST 341

RESULT 3

730337
POLYPROTEIN - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: J30337

R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.

Submitted to the EMBL Data Library, March 1998

A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from XE

A:Reference number: J30337

A:Accession: J30337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1524 <YAN>

A:Cross-references: EMBL:U81290; NID:92981640; PID:92981641; PIDN:AAC24717.1

Query Match 6.3%; Score 168; DB 2; Length 1524;

Best Local Similarity 30.0%; Pred. No. 0.00025;
Matches 45; Conservative 23; Mismatches 66; Indels 14; Gaps 5;

Qy 24 QAEELGDCGHITVSQDSGMTSKNYPGTPNTVCCKITTPKGRKLRLGDLNIESK 83

Db 972 QGPRYGSCEYILTSSEGVIESPNYLGNYPPDLHCQMRILNPAVKYLRVDLEKTEKD 1031

Qy 84 T--CASDYLFS---SADDQYPCGSNAVPEKRLNSNEVYLFKSGSHISGRGFLTY 138

Db 1032 VSGCDDPLLYVNGIGESKDLGVCGE--IRLSISSESEITLPTSNSEVSGKGFSLKY 1090

Qy 139 ASSDHPDLITCLERSHYFEERYSKF-CPA 167.

Db 1091 SFWDK-----QAGSKOLEENAAVGCDA 1113

RESULT 4

108618
Intrinsic factor-B12 receptor CUBILIN precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Aug-2002

C:Accession: T08618

R:Moestrup, S.K.; Kozyrakl, R.; Kristiansen, M.; Kayser, J.H.; Rasmussen, H.H.; Bratt

J. Biol. Chem. 273, 5235-5242, 1998

A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibody

A:Reference number: Z16459; MUID:98148073; PMID:9478979

A:Accession: T08618

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3623 <MOE>

A:Cross-references: EMBL:AF022247; NID:93834379; PIDN:AAC71661.1; PID:93834380

C:Genetics:

A:Gene: CUBILIN

C:Superfamily: intrinsic factor-B12 receptor cubilin; EGF homology

C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-3623/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>

F:133-164/Domain: EGF homology <EGF1>

F:436-467/Domain: EGF homology <EGF2>

Query Match 6.3%; Score 167.5; DB 2; Length 3623;
Best Local Similarity 20.7%; Pred. No. 0.00094;

Matches 110; Conservative 64; Mismatches 189; Indels 169; Gaps 26;

Qy 32 CCHITVSQDSGMTSKNTPGTPNTVCCKITTPKGRKLRLGDLNIE-SKTCASDYL 90

Db 932 CEVTLA-STGILIESPGRNYPVPCVNTWVAVVQGRGILRLEFSSFLIEFYNTNDYL 990

Qy 91 -LEFSATDQY-GPYGSNAVPEKRLNSNEVYLFKSGSHISGRGFLTYASSD----- 142

Db 991 EITDTAAQTFLGRYCGK-SIPSLSNSNSIKLITVSQALAHGEGFSINTEAIDSSVCL 1049

Qy 143 -----HPDLITCLER-----GSHYFEERYSKPCAGCRDIAGD 175

Db 1050 YDITDNFGMLSSPNPNPNYPSWMECIYRLTWGLNQIALHFDPLFEDYFGQCYDFV-- 1107

Qy 176 ISGNTKDGTRDTSLCKAALHAG-IITDELGCHIN-----LLOSKGISHT----- 219

Db 1108 ---EIRDDGYETSPV--VGICYGSVLPPTIISHSKMLKFKSDAALAKGFSAYWDCSS 1162

Qy 220 EGLLAN-----GVL-----SRHGLSKRR----- 239

Db 1163 TGGCGNLTPPTGVLSPNTPMPYHSSSECYWRLESHSGSPFELFQDRLHHPSCSLDY 1222

Qy 240 ---LEFTPGMN-----ITVAIPSVITFAL-----LTGNGTIFATCRKRRKGN 280

Db 1223 LAVFDGPTTNSLIDKLGDTTPAPIRSNKQVYLAKLTADAGQGRGEINRQKCDN-- 1280

Qy 281 PYVSADAKTGCWKQIKYPPARHOSSTFTISYDNKEKMTQKLDLITSPMADYQOPLMTGT 340

Db 1281 --VIVNKTSGILIESINP-----NPYDKNQRCNMTIQAOTGTGMVNV----- 1320

Qy 341 GTVAKSGSTFRMDTDEEVNRTAASGHYDCPHRGRHEVALPLTHSEPEYATPIVERH 400

Db 1321 -----FLLGFVDS-YNMCSTDVYELDGMWGRY-----CGNMPPGATYGSQHL 1367
 QY 401 LIRAHF---FSTOSGYRVPGRPTHKSHSGEPFPAWGAQYQYORPASP 449
 Db 1368 VL-FHFDGINSGEKGRM-----QMFTHCCGGEMSGTAG-----SFSSPPTP 1408

RESULT 5
 T09456
 Intrinsic factor-B12 receptor cubillin precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Aug-2002
 C:Accession: T09456
 R:Koziyaki, R.; Kristiansen, M.; Sliatkaroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N
 Blood 91, 3593-3600, 1998
 A>Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characteriz
 lon.
 A:Reference number: Z16677; M0ID:98241400; PMID:9572993
 A:Accession: T09456
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3623 <KOZ>
 A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AAC82612.1; PID:93929529
 C:Genetics:
 A:Map position: 10p12
 C:Superfamily: Intrinsic factor-B12 receptor cubillin; EGF homology
 C:Keywords: receptor; vitamin B12 uptake
 F:1-24/Domain: signal sequence #status predicted <Sig>
 F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <Mat>
 F:436-467/Domain: EGF homology <EGF>

Query Match
 Best Local Similarity 6.1%; Score 162.5; DB 2; Length 3623;
 Matches 47; Conservative 24; Mismatches 57; Indels 37; Gaps 7;

QY 5 AGPSTALIFAVCAPLR-----LQAEELG-----DCGHITVS 38
 Db 2640 ADGP---LMMRLCGPSKPLPLVPIPSQWIFHTNEREHGFGNAKSYFTDGGIQTG 2695

QY 39 ODGTSKSNPGTYPNMYCEKIIIVPKGRILRLGDNIE-SKTCASDYLLF-----S 93
 Db 2696 -DSGVITSPYPMAYDLSLHCKSWLAPGCHTTLTFSDDIPHTTCAMDVATVRNGGS 2754

QY 94 SATDQYPCYGSWAVPKELRLNSNEVTLFKSGSHISGRGFLTY 138
 Db 2755 PESPIIGYCGN-SNRPRTQSGSNQLVTFNSDHSILGCGGFATW 2798

RESULT 6
 S05008
 Complement subcomponent C 1SBAK.GIF (EC 3.4.21.42) precursor [similarity] - golden hamst
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C:Accession: S05008
 R:Kinoshita, H.; Sakiyama, H.; Tokunaga, K.; Imajob-Ohmi, S.; Hamada, Y.; Isono, K.; Sak
 FEBS Lett. 250, 411-415, 1989
 A>Title: Complete primary structure of a calcium-dependent serine proteinase capable of
 A:Reference number: S05008; M0ID:89325606; PMID:2753140
 A:Accession: S05008
 A:Molecule type: mRNA
 A:Residues: 1-695 <KIN>
 A:Cross-references: EMBL:X16160; NID:949621; PIDN:CA34286.1; PID:949622
 A:Note: part of this sequence, including the amino ends of both the heavy and light cha
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
 C:Keywords: beta-hydroxyasparagine; calcium binding; duplication; glycoprotein; hydrolas
 F:1-21/Domain: signal sequence #status predicted <Sig>
 F:17-133/Domain: C1r/C1s repeat homology <C1R>
 F:122-444/Product: serine proteinase heavy chain #status experimental <HCH>
 F:141-177/Domain: EGF homology <EGF>
 F:181-293/Domain: C1r/C1s repeat homology <C1R2>
 F:300-360/Domain: complement factor H repeat homology <FH1>
 F:365-428/Domain: complement factor H repeat homology <FH2>

F:445-682/Domain: trypsin homology <TRY>
 F:446-695/Product: serine proteinase light chain #status experimental <LCH>
 F:71-89,141-153,149-162,164-177,181-208,240-257,300-347,327-360,365-410,392-428,432-5
 F:155/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:180,413/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:482,536,638/Active site: His, Asp, Ser #status predicted

Query Match
 Best Local Similarity 5.9%; Score 158.5; DB 1; Length 695;
 Matches 48; Conservative 37; Mismatches 76; Indels 25; Gaps 7;

QY 15 FAVCAP---LRLQAEELGDCGHITVSODSGTMTSKNYPGTYPNTVCEKIIIVPKGRIL 71
 Db 161 FCSCPPEYFHDMDRMKGVNCGNVFTALIGETISPPYPPYDENSRCFQILLEGFGY 220

QY 72 ILRL--GDNIE---SKTCASDYLLSSANDQYPCYGS-WAVPKELRLNSNEVTLFKS 125
 Db 221 VVIQREDFVDEPADSGNCKQSDSLFPAKNRQFPGCGNFPGLTLETHTSLDITVFT 280

QY 126 GSHISGRGFLTYASSDHPDLITC-----LEKSHYFEERYSKPCAGCHDIAG 174
 Db 281 DLFEQKKGWKLRY-----HGDPICPKREITANSVWAPDEKATYRKDVYKISCVDFEAVEG 336

QY 175 DISGNT 180
 Db 337 NV-GST 341

RESULT 7
 J00948
 A5 antigen precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: J00948
 R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
 Neuron 7, 295-307, 1991
 A>Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homol
 A:Reference number: J00466; M0ID:91337458; PMID:1908252
 A:Accession: J00466
 A:Molecule type: mRNA
 A:Residues: 1-927 <TAK>
 A:Cross-references: GB:D10467; GB:D01077; NID:9222962; PIDN:BA01360.1; PID:9222963
 A:Experimental source: tadpole, brain
 A:Note: this protein has motifs homologous to complement components C1r and C1s and t
 C:Comment: This protein is a neuronal cell surface molecule involved in the neuronal
 C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-termi
 C:Keywords: duplication; glycoprotein; transmembrane protein
 F:1-21/Domain: signal sequence #status predicted <Sig>
 F:22-927/Product: A5 antigen #status predicted <A5A>
 F:27-138/Domain: C1r/C1s repeat homology <C1R>
 F:147-262/Domain: C1r/C1s repeat homology <C1R2>
 F:274-424/Domain: discoidin I amino-terminal homology <DN1>
 F:446-812/Domain: discoidin I amino-terminal homology <DN2>
 F:561-883/Domain: MAM homology <MAM>
 F:150-261,300,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match
 Best Local Similarity 5.9%; Score 157; DB 1; Length 927;
 Matches 99; Conservative 80; Mismatches 177; Indels 152; Gaps 25;

QY 21 LRLQAEELGDCGHITVSODSGTMTSKNYPGTYPNTVCEKIIIVPKGRILRLDLNT 80
 Db 136 IREYVKTGREGSRNFTS-SNGVIRKPKYDEKXPNMLECTYIIIFARKMGEIVLEFSPFL 194

QY 81 ES-----KTCSADYL---LSSATDQYPCYGSWAVPKELRLNSNEVTLFKSGSHI 129
 Db 195 EADSNAPGGGTCRYDMIGWDGFPVGPPIHGRVCGO-NTPGRVRSFTGLISMIFHDSAI 253

QY 130 SGSGFLTYASDHPDLITCLEKSHYFEERYSKPCAGCHDIAGISGNTKGYVDTSI 189
 Db 254 AKGCFRANFSV-----VQSTEDDFQ----- 274


```

OY 190 LCKAHAGITDELGHINILNLSKISHYEGLLANGVLSRHGSLSEKRFLET---TPG 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 -CKBAL--GMSEGT-----HFDQISVSQYSNMWSERLNLVENGWTPG 318
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 246 MNTTVAIPSVIFALLLTGMI-FAICRRKKK-----GNPVAS-ADAAK- 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 EDVKEWQVDLEMLRFVSGIGTQALSKERKKYFKVSYVDISSNEDMTLLDGKKH 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 290 -----TGCWQKIYKPFARHSTET-----ISYDNEKEMTOKLDTSDMADYQPLMIG 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 LVFNGNDATQVYVRFPEKPYITRFVRLRPVTWEN--GISLRFELYGCKINDYPCSRRLG 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 340 --TGTVAARKGT-FRPMDDT--TEEVRYNTASGHYDCPHRPGRHEVALPLTHSEPE--- 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 437 MVSGLISDSQITASSQVDRNVPBELARLYTSRSQ-----WALPSPNTHPYTKE 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 392 -YANPIYERHLLRAHTFSTQSGYRVGPRPHKS-----HSSGFPFPAATGATOV 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 WLQDLAEKIVR-----GVIIQSGK--HKEKVPFRKFKIGYSNNG-----TEW 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 441 ESYORPASPKPV---GGGYDKPAAASP 464
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 EMIMDSKNKPFTEGNNVYDTPELARTF 555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
N:Reactive factor (EC 3.4.21.-) 1 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
C:Date: 19-May-2000 #sequence, revision 19-May-2000 #text, change 16-Jun-2000
C:Species: Homo sapiens (man)
C:Accession: 154763; JN0883
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
Int. Immunol. 6, 665-669, 1994
A:Title: Molecular characterization of a novel serine protease involved in activation of
A:Reference number: 154763; MUID:94289349; PMID:8018603
A:Accession: 154763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-699 <SAT>
A:Cross-references: GB:D28593; NID:9790963; PIDN:BA05928.1; PID:9471128
R:Tokada, F.; Takayama, Y.; Hatsuue, H.; Kawakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
A:Title: A new member of the C1s family of complement proteins found in a bactericidal
A:Reference number: JN0883; MUID:94059062; PMID:8240317
A:Accession: JN0883
A:Molecule type: mRNA
A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <FAK>
A:Cross-references: DDBJ:D17525; NID:9439712; PIDN:BA04477.1; PID:9439713
A:Experimental source: liver
C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 comp
C:Genetics:
A:Gene: GDB:MASP1; CRAP1; PRSS5; MASP
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521
A:Map position: 3627-3628
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
C:Keywords: beta-hydroxyaspartine; complement pathway; duplication; glycoprotein; hydro
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>
F:19-135/Domain: C1r/C1s repeat homology <C1R1>
F:143-181/Domain: EGF homology <EGF>
F:185-284/Domain: C1r/C1s repeat homology <C1R2>
F:301-362/Domain: complement factor H repeat homology <FH1>
F:367-432/Domain: complement factor H repeat homology <FH2>
F:449-691/Domain: trypsin homology <TRY>
F:478,407/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:73-91,143-157,153-166,168-181,185-212,242-260,301-349,329-362,367-414,337-432,436-572,
F:159/Modified site: erythro-beta-hydroxyaspartine (Asn) #status predicted
F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted
F:450,552,646/Active site: His, Asp, Ser #status predicted

```

```

Matches 48; Conservative 45; Mismatches 89; Indels 27; Gaps 8;
OY 32 CGHITYSODSGTMTSKNPGTYPTVCEKITVPKGRLLIRGLD-LNRSK---TCAS 87
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 CSDNLFQRTQVITSPDPNPNPKSSSECLTYITELLEGVMVLQGFEDIDIDHEVPCPY 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 88 DYLFFSSATDQYGYCGSWAVPKELRLNSNEVTLFSGSHISGRFLTY--ASSDHPD 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 DYIKIKVQPKVIGPFCGKKA-PEPISQSHSVLLIFHSNDNAENRWMLSYRAAGNECPE 303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 146 LITGL-----ERSGHYEEKSKYKFCPCAGCDIAGDISGN-----KGYRDSL-LCK 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 LQPPHKGIEPSQAKYFEPKQDVLVSCDTGYVLDQNVEMDTFQIECLKDGWMSKPIPTCK 363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 193 -----AAIHAGITDELGHINILNLSK 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 IVDCAPELHGLITFSTRNNLTYYKSE 392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
C1HUS
complement subcomponent C1SBAK.GIF (EC 3.4.21.42) precursor [validated] - human
N:Alternate names: C1 esterase precursor
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text, change 03-Jun-2002
C:Accession: A40496; A27381; S00224; S26732; S05634; A05140; A25396; A38407; B37820
R:Kusumoto, H.; Hirosewa, S.; Saller, J.P.; Hagen, F.S.; Kurachi, K.
Proc. Natl. Acad. Sci. U.S.A. 85, 7307-7311, 1988
A:Title: Human genes for complement components C1r and C1s in a close tail-to-tail ar
A:Reference number: A40496; MUID:89017187; PMID:2459702
A:Accession: A40496
A:Molecule type: mRNA
A:Residues: 1-688 <KUS>
A:Cross-references: GB:U04080; NID:9179645; PIDN:AA51852.1; PID:9179646
R:Tosti, M.; Duponchel, C.; Meo, T.; Jullier, C.
Biochemistry 26, 8516-8524, 1987
A:Title: Complete cDNA sequence of human complement C1s and close physical linkage of
A:Reference number: A27381; MUID:86165522; PMID:2831944
A:Accession: A27381
A:Molecule type: mRNA
A:Residues: 1-688 <ROS>
A:Cross-references: GB:M18767; NID:9179647; PIDN:AA51853.1; PID:9179648
R:MacKinnon, C.M.; Carter, P.E.; Smyth, S.J.; Dunbar, B.; Fothergill, J.E.
Eur. J. Biochem. 169, 347-353, 1987
A:Title: Molecular cloning of cDNA for human complement component C1s. The complete a
A:Reference number: S00224; MUID:88082788; PMID:3500856
A:Accession: S00224
A:Molecule type: mRNA
A:Residues: 1-688 <MAC>
A:Cross-references: EMBL:X06596; NID:929542; PIDN:CA29817.1; PID:9763110
A:Accession: S26732
A:Molecule type: protein
A:Residues: 16-38,68-116,170-236,246-262,265-280,282-284,287-308,315-363,384-394,421-
R:Tosti, M.; Duponchel, C.; Meo, T.; Couture-Tosti, E.
J. Mol. Biol. 208, 709-714, 1989
A:Title: Complement genes C1r and C1s feature an intronless serine protease domain cl
A:Reference number: S05634; MUID:90040704; PMID:2553984
A:Accession: S05634
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 356-513, 'G', 514-688 <TO2>
R:Carter, P.E.; Dunbar, B.; Fothergill, J.E.
Biochem. J. 215, 565-571, 1983
A:Title: The serine proteinase chain of human complement component C1s. Cyanogen brom
A:Reference number: A05140; MUID:84104122; PMID:6362661
A:Accession: A05140
A:Molecule type: protein
A:Residues: 438-463, 'X', 485-500, 503-534, 542-558, 561-572, 'A', 574-601, 617-623, 626-644, 6
R:Spycher, S.E.; Nick, H.; Rickli, E.E.
Eur. J. Biochem. 156, 49-57, 1986
A:Title: Human complement component C1s. Partial sequence determination of the heavy
A:Reference number: A25396; MUID:86164350; PMID:3007145
A:Accession: A25396

```

Query Match 5.7%: Score 151.5; DB 1; Length 699;
Best Local Similarity 23.0%: Pred. NO. 0.0015;

A:Molecule type: protein
 A:Residues: 16-61;168-219;287-293,'K',295-334;384-445 <SPY>
 R:Hess, D.; Schallier, J.; Rickli, E.E.
 Biochemistry 30, 2827-2833, 1991
 A:Title: Identification of the disulfide bonds of human complement C1s.
 A:Reference number: A38407; MUID:91175725; PMID:2007122
 A:Accession: A38407
 A:Molecule type: protein
 A:Residues: 131-134,'X',136-146,'X',148-150;155,'X',157-162;166-170,'X',172-174,'X',176-
 'X',387-402,'X',404-408;416-424,'X',426-431;547-556;592-597;617,'X',619-627,'X',629-635
 R:Thielens, N.M.; Aude, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.
 J. Biol. Chem. 265, 14469-14475, 1990
 A:Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated NH-
 A:Reference number: A37820; MUID:90354439; PMID:2387866
 A:Accession: B37820
 A:Molecule type: protein
 A:Residues: 16-25;'X',203-207 <PHI>
 R:Thielens, N.M.; Van Dorsselaer, A.; Gagnon, J.; Arlaud, G.J.
 Biochemistry 29, 3570-3578, 1990
 A:Title: Chemical and functional characterization of a fragment of C1s containing the EF
 A:Reference number: A32672; MUID:90283368; PMID:2141278
 A:Contents: annotation; erythro-beta-hydroxyasparagine site, content
 A:Note: about half of the A chains contain erythro-beta-hydroxyasparagine
 C:Comment: This protein is a serine protease that combines with C1q and C1r to form C1
 s C2 and C4.
 C:Comment: C1s is a dimer of identical chains, each of which is activated by cleavage in
 C:Genetics:
 A:Gene: GDB:C1S
 A:Cross-references: GDB:119730; OMIM:120580
 A:Map position: 12p13-12p13
 A:Introns: 291/1; 329/3; 356/1; 399/1; 424/1
 A:Note: the list of introns may be incomplete
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
 C:Keywords: acute phase; beta-hydroxyasparagine; calcium binding; complement pathway; du
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:11-127/Domain: C1r/C1s repeat homology <C1R1>
 F:16-688/Product: complement subcomponent C1s #status experimental <MAT>
 F:16-437/Product: complement subcomponent C1s chain A (heavy chain) #status experimental
 F:135-171/Domain: EGF homology <EGF>
 F:175-287/Domain: C1r/C1s repeat homology <C1R2>
 F:294-354/Domain: complement factor H repeat homology <RH2>
 F:359-421/Domain: complement factor H repeat homology <RH1>
 F:438-688/Product: complement subcomponent C1s chain B (light chain) #status experimental
 F:438-679/Domain: trypsin homology <TRY>
 F:65-83,'X',135-147,'X',156-158-171,'X',175-202;234-251;294-341;321-354;359-403;386-421;425-549,
 F:149/Modified site: erythro-beta-hydroxyasparagine (Asn) (partial) #status experimental
 F:174,406/Blinding site: carboxylate (Asn) (covalent) #status experimental
 F:437-438/Cleavage site: Arg-Ile (complement subcomponent C1r) #status experimental
 F:475,529,632/Active site: His, Asp, Ser #status predicted

Query Match 5.7%; Score 151; DB 1; Length 688;
 Best Local Similarity 25.3%; Pred. No. 0.0017;
 Matches 46; Conservative 31; Mismatches 89; Indels 16; Gaps 5;

QY 15 FAVCAP---LRQAELGDCGHTVVSODSGTSTKNTPTVNTVCEKLTTPPKGKL 71
 DB 155 FSCSPETVFLHDMKNCVNSGDVFTALIGELASPNTPYPPENSRCEYQILERGFQV 214
 QY 72 I--LRGLDNIESTKTCAS---DYLLFSSATDQYPCYCG--SWAPKRELRLNSENVLERS 125
 DB 215 VVTLREDDVDEPAADSAKNCILSLVAVAGDRGCPYCGHFPPLNIETKSNLIDLIFQF 274
 QY 126 GSHISGRGLTY-----ASSDHPDLITCLERGSHYEERKSKPCPACRDIADISG 178
 DB 275 DLTGKRGKMLRYHGDPMPCKEDTPNSWEPKAKYVFDVVOITCLDGFVEVEGVGA 334
 QY 179 NT 180
 DB 335 TS 336

hensin - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30549
 R:Takito, J.; Hikita, C.; Al-Awqati, Q.
 J. Clin. Invest. 98, 2324-2331, 1996
 A:Title: Hensin, a new collecting duct protein involved in the in vitro plasticity of
 A:Reference number: Z20851; MUID:97096804; PMID:8941650
 A:Accession: T30549
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-1594 <TAK>
 A:Cross-references: EMBL:AF043112; MUID:94105083; PID:94105084; PIDN:AAD02242.1

Query Match 5.6%; Score 150; DB 2; Length 1594;
 Best Local Similarity 25.6%; Pred. No. 0.0065;
 Matches 65; Conservative 34; Mismatches 105; Indels 50; Gaps 12;

QY 32 CGHIVTSQDSGTMTSKNPGTYVYCEKLTTPPKGRLILRLGDLNIESKT-CASDYL 90
 DB 951 CGGFLVNA-TGSPSPSPYPYPPNNAICVWEIAPSGYLINLAFSOLRLQHSVCNEDY 1009
 QY 91 -LSSARDQ--YPRYGSMAVPRKELRLNSENVTYFRSSGSHISGRGLTYAS----- 140
 DB 1010 EIRPDGSDSSLLKRICNDSC--QIFTTSSNRMTVLFPSDLSVONTGLFVANSFPDASL 1067
 QY 141 ---SDHPLITCLERGSHYEERKSKPCPACRDIADISGNTKDG-YRDTSLCKRAIH 196
 DB 1068 RLYSGNSYACAGREIYHIGRWYVC-----DSWDYQDAQVYCRQJCGDAVS 1118
 QY 197 A-----GIIT-DEL---GGHINLQSKGISHYEGLANGVLSRRGSLSEKRLFT 242
 DB 1119 APGAVFGSGSGGTLTLDVNCSTGTEATLMQCRQSWP-----SHNGCHHEDASVIC 1169
 QY 243 TPQMGNTITVAIPSY 256
 DB 1170 TGNVGTTSASVPI 1183

RESULT 11
 CHDRB
 complement subcomponent C1RBAR.GIF (EC 3.4.21.41) precursor [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 15-Nov-1984 #sequence_revision 30-Jun-1991 #text_change 03-Jun-2002
 R:Leysen, A.; Kurachi, K.; Sakuribayashi, K.S.; Davie, E.W.
 Biochemistry 25, 4855-4863, 1986
 A:Title: Nucleotide sequence of the cDNA coding for human complement C1r.
 A:Reference number: A24170; MUID:87026366; PMID:3021205
 A:Accession: A24170
 A:Molecule type: mRNA
 A:Residues: 1-705 <LEY>
 A:Cross-references: GB:M4058; NID:q179643; PIDN:AAA51851.1; PID:q179644
 R:Journet, A.; Tosi, M.
 Biochem. J. 240, 783-787, 1986
 A:Title: Cloning and sequencing of full-length cDNA encoding the precursor of human c
 A:Reference number: A29768; MUID:87156625; PMID:3030286
 A:Accession: A29768
 A:Molecule type: mRNA
 A:Residues: 1-151,'U',153-705 <JOU>
 A:Cross-references: GB:X04701; NID:299538; PIDN:CAA28407.1; PID:299539
 R:Arlaud, G.J.; Willis, A.C.; Gagnon, J.
 Biochem. J. 241, 711-720, 1987
 A:Title: Complete amino acid sequence of the A chain of human complement-classical-pa
 A:Reference number: A29769; MUID:87241248; PMID:3036070
 A:Accession: A29769
 A:Molecule type: protein
 A:Residues: 18-166,'X',168-463 <ARL>
 A:Note: 152-Leu was also found
 R:Arlaud, G.J.; Van Dorsselaer, A.; Bell, A.; Mancini, M.; Aude, C.; Gagnon, J.
 FEBS Lett. 222, 129-134, 1987
 A:Title: Identification of erythro-beta-hydroxyasparagine in the EGF-like domain of h
 A:Reference number: S02422; MUID:88005128; PMID:2820791

A:Molecule type: mRNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:I.24755; NID:q439606; PIDN:AAA37306.1; PID:q439607
C:Genetics:
A:Gene: Bmp-1
C:Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology; F:
F:135-326/Domain: astacin homology <AST>
F:556-592/Domain: EGF homology <EGF>
F:596-705/Domain: C1r/C1s repeat homology <C1R>
F:712-747/Domain: EGF homology <EG2>
F:218,222,228,277/Binding site: zinc (His, His, His, Tyr) #status predicted
F:219/Active site: Glu #status predicted

Query Match 5.3%; Score 142.5; DB 2; Length 991;
Best Local Similarity 33.0%; Pred. No. 0.013;
Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;

QY 31 GCGHIVTSODSGTMTSKNYPGTYNYVCEKIIIVPKGRLLRLGLDNLIESK--TCASDY 89
DB 751 GCEHRTVS--TSGRTSPNMDPKYPSKKECTWAIISTPGHRYKLTFFVEMDIESQECAYDH 809
QY 90 LFFSSATDQYGR---YCGSWAVPKELRLNSNEYVLEKSGSHISGRGFLITYAS 140
DB 810 LEVFDGRDAKAPVIGRFCS--KRPEVLAIGNRFLEFSDNSVQRKGFQASHST 863

RESULT 14

JC7629

membrane-type frizzled-related protein - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7629
R:Kato, M.
Biochem. Biophys. Res. Commun. 282, 116-123, 2001

A:title: Molecule cloning and characterization of MFRP, a novel gene encoding a membra
A:Reference number: JC7629; MUID:21164708; PMID:11263980
A:Accession: JC7629
A:Molecule type: mRNA
A:Residues: 1-579 <RAT>
A:Cross-references: DDBJ:AB055505
C:Comment: This protein, which plays key roles in medulla oblongata as a regulator of th
ubulin domains.
C:Genetics:
A:Gene: mfrp
A:Map position: 11q23
C:Keywords: transmembrane protein

Query Match 5.3%; Score 142; DB 2; Length 579;
Best Local Similarity 23.7%; Pred. No. 0.0064;
Matches 58; Conservative 37; Mismatches 88; Indels 62; Gaps 10;

QY 31 GCGHIVTSODSGTMTSKNYPGTYNYVCEKIIIVPKGRLLRLGLDNLIESK--TCASDY 89
DB 300 GCGGNLTGL--GTFSTPSTYLAQYFHOILCTWHLISVPAGHSTELDFHNFSLERODECKFDY 358
QY 90 L-----LFSATDQYGRPGYCSWAVPKELRLNSNEYVLEKSGSHISGRGFLITYASD-- 142
DB 359 VEYETSSGAFSLGRFCGA--EPPPHLVSHHSLAVLFRDHISSGFSATYLAFNAT 417
QY 143 ----HPLITCLERGSYFEKYSKFCPCAGCDIAG--DISGNTKDYRDT----- 187
DB 418 ENPCGPSELSCQ-----AGCGKGVQMMCDMWRDCTGSDSDNCSGPLFPP 462
QY 188 SLICKAA-----IHAGITDE-----LGGHINLQSKGISHEGLLAN 225
DB 463 ELACEPVQVEMCLGSLNTTAFTPIWGMITIQEEVVEVLSGYKSLTSLPTQHFRLLCG 522
QY 226 GVLSR 230
DB 523 LTVPR 527

RESULT 15

T42721

CRP-ductin-alpha precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 23-Mar-2001
C:Accession: T42721
R:Cheng, B.; Bjerknes, M.; Chen, H.
Anal. Rec. 244, 327-343, 1996

A:title: CRP-ductin: a gene expressed in intestinal crypts and in pancreatic and hepa
A:Reference number: 222241; MUID:96362470; PMID:8742698
A:Accession: T42721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2083 <CHE>
A:Cross-references: EMBL:U37438; NID:q1276646; PID:q1276647; PIDN:AAC52505.1
A:Experimental source: strain BALB/c; jejunal epithelial cells
C:Keywords: transmembrane protein
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2083/Product: CRP-ductin-alpha #status predicted <MAT>

Query Match 5.3%; Score 142; DB 2; Length 2083;
Best Local Similarity 24.1%; Pred. No. 0.04;
Matches 70; Conservative 39; Mismatches 91; Indels 90; Gaps 16;

QY 32 GCHIVTSODSGTMTSKNYPGTYNYVCEKIIIVPKGRLLRLGLDNLIESKTCASDY- 90
DB 1631 GCGFLT--QPSGQFSSPFYPGNYPNMARCLMNIIEVPNNRVTVPRDQLE--RGCSYDYE 1688
QY 91 LFSATDQYGRPGYCSWAVPK-----ELRLNSNEYVLEKSGSHISGRGFLITYASDH 143
DB 1689 IFD-----GPHSSPLIARVCDGLSFGSTNFMMSIRFTDHSIYARGP----- 1733
QY 144 PDLITCLERGSYFEKYSKFCPCAGCDIAGDISGNTKDYRDTSLICKAIIHAGITDE 203
DB 1734 -----QAHY-----SDPNNNT-----TNLQGS----- 1752
QY 204 LGCHI-----NLQSGISHYEGILANGVLSRH--GSLSEKRLFTTGMNITTV---- 251
DB 1753 --NHQASVSRSYLSQMGYSARDLVIPGMNSYHCQPOITQREVIFETIPYGCCTIKOAD 1810
QY 252 --ALPSYIFALLLTGMGIFALCRKKRKKGPYASD--AOKTGCKWKQIKY 298
DB 1811 NETINYSFLRAVVSN--GITI---KRRRDINIHVSCRLQNT--WVNTMY 1853

Search completed: May 15, 2003, 13:17:00
Job time : 23.454 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 13.6281 Seconds

(Without alignments)
1530.854 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPGVLAALFAVCAV.....YSAPRNGLAFLNQTAAATALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181.5	6.8	686	1	MAS2_HUMAN
2	158.5	5.9	695	1	CASP_MESOC
3	157.5	5.9	704	1	CRAR_MOUSE
4	151.5	5.7	928	1	NRP1_XENLA
5	151.5	5.7	699	1	CRAR_HUMAN
6	151.5	5.7	688	1	CIS_HUMAN
7	147	5.5	705	1	CIR_HUMAN
8	145	5.4	550	1	COCH_HUMAN
9	143.5	5.4	923	1	NRP1_MOUSE
10	142.5	5.3	991	1	BMP1_MOUSE
11	141.5	5.3	986	1	BMP1_HUMAN
12	141	5.3	514	1	UVS2_XENLA
13	141	5.3	639	1	BMP6_STRPU
14	140.5	5.3	468	1	PCOI_RAT
15	139	5.2	1022	1	TLID_BRARE
16	135	5.1	449	1	PCOI_HUMAN
17	133.5	5.0	468	1	NRP2_MOUSE
18	132.5	5.0	931	1	NRP2_HUMAN
19	132	4.9	547	1	COCH_CHICK
20	131.5	4.9	855	1	STI4_HUMAN
21	131.5	4.9	922	1	NRP1_RAT
22	131.5	4.9	925	1	NRP2_RAT
23	130	4.9	552	1	COCH_MOUSE
24	129.5	4.8	707	1	NRP2_MOUSE
25	127.5	4.8	707	1	NRP1_XENLA
26	127	4.8	855	1	STI4_MOUSE
27	125	4.7	1019	1	LFC_MOUSE
28	124	4.6	1057	1	TLID_DROME
29	122	4.6	616	1	SPAN_STRPU
30	121.5	4.5	923	1	NRP1_HUMAN
31	121	4.5	1019	1	LFC_CARRO
32	120.5	4.5	597	1	BPIO_PARLI
33	119.5	4.5	914	1	NRP1_CHICK

34	116	4.3	275	1	TSG6_MOUSE	O08859	mus musculus
35	115	4.3	277	1	TSG6_HUMAN	P98066	homo sapien
36	113.5	4.2	1064	1	FBP1_STRPU	P10079	strongyloce
37	110	4.1	276	1	TSG6_RABIT	P98065	oryctolagus
38	110	4.1	1369	1	MSH5_CAEEL	Q19272	caenorhabdi
39	105.5	3.9	1493	1	NEO1_MOUSE	P97798	mus musculus
40	105	3.9	620	1	EXTN1_TOBAC	P13963	nicotiana t
41	105	3.9	968	1	BCAL_RAT	O63767	rattus norv
42	104.5	3.9	570	1	FBP3_STRPU	P49013	strongyloce
43	104.5	3.9	1069	1	ENTR_MOUSE	P97435	mus musculus
44	104.5	3.9	1461	1	NEO1_HUMAN	Q92859	homo sapien
45	104	3.9	310	1	ASTL_CORJVA	P42662	colurnix co

ALIGNMENTS

RESULT 1
MAS2_HUMAN STANDARD; PRT; 686 AA.
ID MAS2_HUMAN
AC 000187; 075754; 09Y270; 09B2H0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)
DE (Mannose-binding protein associated serine protease 2) (MASP-2)
DE (MBL-associated serine protease 2).
GN MASP2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97242412; PubMed=9087411;
RA Thiel S., Jensen T.V., Stover C.M., Schwaible W.J., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "A second serine protease associated with mannan-binding lectin that
RT activates complement.";
RL Nature 386:506-510(1997).
[2]
RP SEQUENCE FROM N.A.
RX Thiel S., Vorup-Jensen T., Stover C.M., Schwaible W., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "Identification and characterization of a novel protein of the human
RT complement system, mannan-binding lectin-associated serine protease-2
RT (MASP-2).";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99192764; PubMed=10092804;
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
RA Jensenius J.C., Schwaible W.J.;
RT "Two constituents of the initiation complex of the mannan-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene.";
RL J. Immunol. 162:3481-3490(1999).
[4]
RP SEQUENCE FROM N.A.
RX Park D., Kim B., Baek K., Yoon J.;
RT "Structure of human MASP-2 gene.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP FUNCTION: TRYPSIN PROTEASE THAT PROBABLY PLAYS AN IMPORTANT ROLE
RP IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER
RP ACTIVATION IT CLEAVES C4 GENERATING C4A AND C4B.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 BGF-LIKE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

 DR EMBL; Y09926; CAA71059.1; -
 DR EMBL; X98400; CAA67050.1; -
 DR EMBL; Y18287; CAB50735.1; -
 DR EMBL; Y18286; CAB50733.1; -
 DR EMBL; AF321562; AAG50274.1; -
 DR EMBL; AF321558; AAG50274.1; JOINED.
 DR EMBL; AF321559; AAG50274.1; JOINED.
 DR EMBL; AF321560; AAG50274.1; JOINED.
 DR EMBL; AF321561; AAG50274.1; JOINED.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.229; -
 DR Genew; HGNC:6902; MASP2.
 DR MIM; 605102; -
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001254; Ser-protease_Try.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00431; CUB; 2.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00032; CCP; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydroxylase; Complement pathway; Serine protease; Protease;
 KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
 FT SIGNAL 1 15
 FT CHAIN 16 686
 FT CHAIN 16 444
 FT CHAIN 445 686
 FT DOMAIN 16 137
 FT DOMAIN 138 181
 FT DOMAIN 184 296
 FT DOMAIN 299 362
 FT DOMAIN 365 431
 FT DOMAIN 445 686
 FT ACT_SITE 483 483
 FT ACT_SITE 532 532
 FT ACT_SITE 633 633
 FT MOD_RES 158 158
 FT DISULFID 72 90
 FT DISULFID 142 156
 FT DISULFID 152 165
 FT DISULFID 167 180
 FT DISULFID 184 211
 FT DISULFID 241 259
 FT DISULFID 300 348
 FT DISULFID 328 361
 FT DISULFID 366 412
 FT DISULFID 396 430
 FT DISULFID 434 552
 FT DISULFID 598 618

FT DISULFID 629 660 POTENTIAL.
 FT CONFLICT 361 362 MISSING (IN REF. 3).
 FT CONFLICT 371 371 D -> Y (IN REF. 4).
 FT CONFLICT 372 372 L -> LGS (IN REF. 3).
 FT CONFLICT 442 442 G -> E (IN REF. 4).
 FT CONFLICT 447 447 G -> E (IN REF. 4).
 FT CONFLICT 461 462 MISSING (IN REF. 3).
 FT CONFLICT 473 473 L -> LIL (IN REF. 3).
 SQ SEQUENCE 686 AA; 75685 MW; 4E34DED159448A2A CRC64;
 Query Match 6.8%; Score 181.5; DB 1; Length 686;
 Best Local Similarity 24.4%; Pred. No. 2.5e-06;
 Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;
 QY 32 CGHYTSQDSGTMTSKNPGYVNTYCEKITVYKGRLLI---RLGDLNIESKT-CAS 87
 DB 184 CSGQVFTQSGSLSSPEYDRPKLSCTYSLSLEGSVILDFEESPDEVETHPPTLCY 243
 QY 88 DYLLSSATDYGPRGCSNAPKELRLNSVYLF---KSGSHISGRGFLITYASSDH- 143
 DB 244 DFLKIQTDRREHGPCCGK-TLPHRIETKSNVTITVFVDESGDH---TGWKIHYSTAH 299
 QY 144 -----PDLITCLBRGSHYFEKYSKCPACGRDIAGDISNT-----KGYRDTSL 190
 DB 300 CYPAPFNGHVSPOAKYTIKDSIFCEYETELQGLPLKSFYAVCQKDSWDRPMP 359
 QY 191 CKAHAGITDELGSHNLQSKGISHYGL 223
 DB 360 ACSIVDCGPPDLLPSGRVEYITGPEVTTYKAVI 392
 RESULT 2
 CASP_MESAU STANDARD; PRT; 695 AA.
 ID CASP_MESAU
 AC P15156;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calcium-dependent serine proteinase precursor (EC 3.4.21.-) (CASP).
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX Mesocricetus.
 RX NCBI_Taxid:10036;
 RA [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-50 AND 446-472.
 RC TISSUE=Embryo;
 RX MEDLINE=89325606; PubMed=2753140.
 RA Kinoshita H., Sakiyama H., Tokunaga K., Imajoh-Ohmi S., Hamada Y.,
 RA Isono K., Sakiyama S.;
 RT "Complete primary structure of calcium-dependent serine proteinase
 RT capable of degrading extracellular matrix proteins.";
 RL FEBS Lett. 250:411-415(1989).
 CC -1- FUNCTION: CAPABLE OF DEGRADING EXTRACELLULAR MATRIX PROTEINS.
 CC CASP DEGRADABLE TYPE I AND IV COLLAGEN AND FIBRONECTIN IN THE
 CC PRESENCE OF CALCIUM.
 CC -1- SUBUNIT: HETERODIMER, CONSISTING OF HEAVY AND LIGHT CHAINS WITH
 CC DISULFIDE BONDS. THE HEAVY CHAIN IS EXPECTED TO BE A REGULATORY
 CC SUBUNIT AND THE LIGHT CHAIN CONTAINS THE CATALYTIC SITE.
 CC -1- DOMAIN: THE GLU-RICH REGION IN THE N-TERMINAL REGION MAY BE GAMMA
 CC CARBOXYLATED AND FUNCTION AS A CALCIUM-BINDING SITE.
 CC -1- SIMILARITY: TO BLOOD COAGULATION FACTORS SUCH AS IX, X AND AN
 CC ANTICOAGULATION FACTOR, PROTEIN C.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Query Match	Best Local Similarity	Score	DB 1	Length	704
Matches 58; Conservative	53; Mismatches 93; Indels 37; Gaps 12	5.9%; Pred. No. 0.00019;			
32	CGHAYTSDSGMTGKNTKPPYNTVVCCKITLVPRGKRLRLRLGLD-LNIESK---TCAS 87				
190	CSGNLFQTGRITGTPDYDNPYPRKSECCYITLDLEGFWMSLQFEDIPDIEHDPEVPKCY 249				
88	DYLFSSATDQYPCYCGSMAVPEKRLNLSNEVTVLFKSGSHISRGFELLTVASSDH--- 143				
250	DYIKAKSGKVMGPPCGEKS-PEPISTGYHVOILFRSDNSGENRGMRLSYRAAGNECPK 308				
144	--POLITGLE--RGSHPYFEKSKSCPKPGCCRIADGISNT-----KQGYRDTSL-LCK 192				
309	LQPYTGKLEPSQAVYSFSDQVAVSCDTGYKVLKONGVMDTQIETCLKDGAWSNKPCTCK 368				
193	-----AIHAGITVDELGCHINILNLSKGIH-----YEGILLAN--GV--LSRHGSLSE 236				
369	IYDCAAPGAKLKGIVLTFSPRNNLTYTKSP-IKYSOOQPYTKMLHNHTGYTCSAHTGTWIN 427				
237	K 237				

RESULT 4	ID	NAME	STANDARD;	PRT;	928 AA.
NRPI_XENIA	NRPI_XENIA	STANDARD;	PRT;	928 AA.	
P28824;					
01-DEC-1992 (Rel. 24, Created)					
01-DEC-1992 (Rel. 24, Last sequence update)					
16-OCT-2001 (Rel. 40, Last annotation update)					
Neuropilin-1 precursor (A5 protein) (A5 antigen).					
Xenopus laevis (African clawed frog).					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;					
Xenopodinae; Xenopus.					
NCBI_TaxID=8355;					
(1)					
SEQUENCE FROM N.A.					
TISSUE=Brain:					
MEDLINE=91337458; PubMed=1908252;					
Taagui S., Hirata T., Agata K., Mochizuki M., Eguchi G., Fujisawa H.,					
"The A5 antigen, a candidate for the neuronal recognition molecule,					
has homologies to complement components and coagulation factors.";					
Neuron 7:295-307(1991).					
-1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE					
CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF					
CEREBRAL NEURONAL CIRCUITS AND IN ORGANOREGULATION OF THE NERVOUS					
SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY					
SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION					
BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.					
-1- SUBCELLULAR LOCATION: Type I membrane protein.					
-1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER					
NEURONS.					
-1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.					
-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.					
-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.					
-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.					

THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION					
BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION					
AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON					
ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY					
MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL					
ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/					
OR SEND AN EMAIL TO license@lsb-sib.ch).					

EMBL; D10467; BAA01260.1; -.					
HSSP; P12259; IC2T.					
InterPro; IPR000859; CUB_domain.					
InterPro; IPR000421; FA58_C.					
InterPro; IPR000998; MAM_domain.					
Pfam; PF00431; CUB; 2.					
Pfam; PF00629; MAM; 1.					
Pfam; PF00734; F5_F8_Type_C; 2.					
PRINTS; PR00020; MAMDOMAIN.					
SMART; SM00042; CUB; 2.					
SMART; SM00231; FA58C; 2.					
SMART; SM00137; MAM; 1.					
PROSITE; PS00740; MAM_1; 1.					
PROSITE; PS01180; CUB; 2.					
PROSITE; PS01285; FA58C_1; 2.					
PROSITE; PS01286; FA58C_2; 2.					
PROSITE; PS00607; MAM_2; 1.					
Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;					
Antigen.					
FT SIGNAL	1	21	POTENTIAL.		
FT CHAIN	22	928	NEUROFILIN-1.		
FT DOMAIN	22	860	EXTRACELLULAR (POTENTIAL).		
FT TRANSMEM	861	883	POTENTIAL.		
FT DOMAIN	884	928	CYTOPLASMIC (POTENTIAL).		
FT DOMAIN	27	141	CUB 1.		
FT DOMAIN	147	265	CUB 2.		


```

FT DOMAIN 275 424 F5/8 TYPE C 1.
FT DOMAIN 431 584 F5/8 TYPE C 2.
FT DOMAIN 646 812 MAM.
FT DISULFID 27 54 PROBABLE.
FT DISULFID 82 104 PROBABLE.
FT DISULFID 147 173 PROBABLE.
FT DISULFID 206 228 PROBABLE.
FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 928 AA; 103416 MW; AF68323B0A4C789D CRC64;

Query Match 5.94; Score 157; DB 1; Length 928;
Best Local Similarity 19.58; Pred. No. 0.0003;
Matches 99; Conservative 80; Mismatches 177; Indels 152; Gaps 25;

OY 21 LRLQAEELGDCGCHIVTSODSGTWSKNPNYTCERTITVPKGRLLIRGLDNI 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 IRYEVFTGPGCSNFPIS-SNGVIKSPKRYPNALACTYIIRAPKQELVLEFESL 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 81 ES-----KTCASDYI---LSSATDQYGPYCGSMAPRELNSNEVTLEKSGSHI 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 195 EADSNAPGCGTCRDYDGLIMDGFPGVPHIGRYCGQ--NTPGRVASFGLISMIFHTDSAI 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 130 SGREGELTVASSDHPDILITCLERSHFEERKSFPCACGCDIAGDISGNTRKDYRDTSL 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 254 AKESGFANFSV-----VOSNTDEDFQ--- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 190 LCKRAHAGITDELGCHINILLOSIGISHYGLANGVLNHSGLSEKRELETT---TPG 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 275 -CKEAL--GHESGRI-----HFDQISVSQSYSMNMAESRLNLYVNGWTPG 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 246 MNTTVAIPSVIFIALILTMKGIT-FAICRKRKR-----GNPYVS-ADAK- 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 319 EDYVKEMIOVDLENRLRVSIGTGAISKEKKKVFVSKYKVDISSNGEDWITDKENKH 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 290 -----TGCWKQIKYPPARHOSFEFT---ISYDNEKEMTOKLDTISDMADYOQPLMIG 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 379 LVFTGNDADVDYRPSKRYITFEVRLRPWTEN--GISLRELYGCKITIDYCSRMIG 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 340 --TGTVAARKST-FRPMDT--TEEVNTEASGHYDCPHRNGHEVALPLTHSEPE--- 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 437 MVSGLISDQITASSQYDRMVMVELARLYTSRSG-----WALPSPNHPYTK 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 392 -YATPIYERILLRAHTSTOSGYVPPRPPIHKS-----HSSGPPPATGATOV 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 485 WLQDLAEKIVR-----GVYIOGCK--HKENKVFMRKFKIGYSNNG-----TEW 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 441 ESYORPASPKPV-----GGYDKPAASSF 464
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 528 EMIMDSKKNPKTEEGNTNTDTPELRTF 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
CRAR HUMAN STANDARD: PRT: 699 AA.
AC P48740; O95570; Q9UE09;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement-activating component of Ra-reactive factor precursor
DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (BARF)
DE (Mannan-binding lectin serine protease 1) (Mannose-binding protein
DE associated serine protease) (MASP-1).
GN MASP1 OR CRARF OR CRARF1 OR PRSS5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Liver;
RX MEDLINE-94059062; PubMed-8240317;
RA Takada F., Takayama Y., Hatsuse H., Kawakami M.;
RT "A new member of the C1s family of complement proteins found in a
RT bactericidal factor, Ra-reactive factor, in human serum.";
RL Biochem. Biophys. Res. Commun. 196;1003-1009(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Fetal liver;
RX MEDLINE-94289349; PubMed-8018603;
RA Sato T., Endo Y., Matsushita M., Fujita T.;
RT "Molecular characterization of a novel serine protease involved in
RT activation of the complement system by mannose-binding protein.";
RL Int. Immunol. 6:665-669(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-97079701; PubMed-8921412;
RA Endo Y., Sato T., Matsushita M., Fujita T.;
RT "Exon structure of the gene encoding the human mannose-binding
RT protein associated serine protease light chain: comparison with
RT complement C1r and C1s genes.";
RL Int. Immunol. 8:1355-1358(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-99402590; PubMed-10475605;
RA Takayama Y., Takada F., Nowatari M., Kawakami M., Matsui-ura N.;
RT "Gene structure of the p100 serine-protease component of the human Ra-
RT reactive factor.";
RL Mol. Immunol. 36:505-514(1999).
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR BARF
CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CC CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1- SUBUNIT: BARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
CC (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)
CC LINKED BY A DISULFIDE BOND.
CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BGE-LIKE DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D17525; BAA04477.1; -
DR EMBL: D28593; BAA05928.1; -
DR EMBL: D61695; BAA34864.1; -
DR EMBL: AB010822; BAA34864.1; JOINED.
DR EMBL: AB010813; BAA34864.1; JOINED.
DR EMBL: AB010814; BAA34864.1; JOINED.
DR EMBL: AB010815; BAA34864.1; JOINED.
DR EMBL: AB010816; BAA34864.1; JOINED.
DR EMBL: AB010817; BAA34864.1; JOINED.
DR EMBL: AB010818; BAA34864.1; JOINED.
DR EMBL: AB010819; BAA34864.1; JOINED.
DR EMBL: AB010820; BAA34864.1; JOINED.
DR EMBL: AB010821; BAA34864.1; JOINED.
DR EMBL: D61690; BAA34864.1; JOINED.
DR EMBL: D61691; BAA34864.1; JOINED.
DR EMBL: D61692; BAA34864.1; JOINED.
DR EMBL: D61693; BAA34864.1; JOINED.
DR EMBL: D61694; BAA34864.1; JOINED.

```

DR	EMBL	AB007617	BAA89206.1	JOINED.	
DR	EMBL	AB007602	BAA89206.1	JOINED.	
DR	EMBL	AB007603	BAA89206.1	JOINED.	
DR	EMBL	AB007604	BAA89206.1	JOINED.	
DR	EMBL	AB007605	BAA89206.1	JOINED.	
DR	EMBL	AB007606	BAA89206.1	JOINED.	
DR	EMBL	AB007607	BAA89206.1	JOINED.	
DR	EMBL	AB007608	BAA89206.1	JOINED.	
DR	EMBL	AB007609	BAA89206.1	JOINED.	
DR	EMBL	AB007610	BAA89206.1	JOINED.	
DR	EMBL	AB007611	BAA89206.1	JOINED.	
DR	EMBL	AB007612	BAA89206.1	JOINED.	
DR	EMBL	AB007613	BAA89206.1	JOINED.	
DR	EMBL	AB007614	BAA89206.1	JOINED.	
DR	EMBL	AB007615	BAA89206.1	JOINED.	
DR	HSSP	P00736	1APO.		
DR	MEROPS	S01.198	-		
DR	Genew	HGNC:6901	MASP1.		
DR	MIM	600521	-		
DR	InterPro	IPR000152	Asx_hydroxyl.		
DR	InterPro	IPR000859	Cub domain.		
DR	InterPro	IPR001314	Chymotrypsin.		
DR	InterPro	IPR000561	EGF-like.		
DR	InterPro	IPR001881	EGF_Ca.		
DR	InterPro	IPR001254	Set_protease_Try.		
DR	Pfam	PFO0084	sush1_2.		
DR	Pfam	PFO0089	trypsin_1.		
DR	Pfam	PFO0431	CUB_2.		
DR	PRINTS	PRO0722	CHYMOTRYPSIN.		
DR	SMART	SM00032	CCP_2.		
DR	SMART	SM00042	CUB_2.		
DR	SMART	SM00181	EGF_1.		
DR	SMART	SM00179	EGF_Ca_1.		
DR	SMART	SM00020	TRY_P_Spc_1.		
DR	PROSITE	PS00010	ASX_HYDROXYL_1.		
DR	PROSITE	PS01180	CUB_2.		
DR	PROSITE	PS01186	EGF_2_1.		
DR	PROSITE	PS01187	EGF_Ca_1.		
DR	PROSITE	PS50240	TRYPSIN_DOM_1.		
DR	PROSITE	PS00134	TRYPSIN_HIS_1.		
DR	PROSITE	PS00135	TRYPSIN_SRR_1.		
KM	Hydrolase	Complement pathway: Serine protease; Protease;			
KM	Glycoprotein	Sush1	Repeat:		
FT	CHAIN	1	19	POTENTIAL.	
FT	CHAIN	20	699	COMPLEMENT-ACTIVATING COMPONENT OF	
FT	CHAIN	20	448	RA-REACTIVE FACTOR.	
FT	CHAIN	449	699	70 KDA CHAIN OF P100 (P70).	
FT	DOMAIN	20	138	29 KDA CHAIN OF P100 (P29).	
FT	DOMAIN	139	182	CUB 1.	
FT	DOMAIN	185	237	EGF-LIKE, CALCIUM-BINDING (POTENTIAL).	
FT	DOMAIN	300	363	CUB 2.	
FT	DOMAIN	366	433	SUSH1 1.	
FT	DOMAIN	449	699	SUSH1 2.	
FT	DOMAIN	490	480	SERINE PROTEASE.	
FT	ACT_SITE	552	552	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	ACT_SITE	646	646	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	MOD_RES	159	159	CHARGE RELAY SYSTEM (BY SIMILARITY).	
FT	DISULFID	73	91	HYDROXYLATION (POTENTIAL).	
FT	DISULFID	143	157	POTENTIAL.	
FT	DISULFID	153	166	POTENTIAL.	
FT	DISULFID	168	181	POTENTIAL.	
FT	DISULFID	185	212	POTENTIAL.	
FT	DISULFID	242	260	POTENTIAL.	
FT	DISULFID	301	349	POTENTIAL.	
FT	DISULFID	329	362	POTENTIAL.	
FT	DISULFID	367	414	POTENTIAL.	
FT	DISULFID	397	432	POTENTIAL.	
FT	DISULFID	436	572	INTERCHAIN (POTENTIAL).	
FT	DISULFID	614	631	POTENTIAL.	
FT	DISULFID	642	672	POTENTIAL.	

FT	CARBONAD	49	49	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOND	178	178	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHD	385	385	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOH	407	407	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	235	235	O -> E (IN REF. 1 AND 4).	
FT	CONFLICT	285	285	A -> G (IN REF. 1).	
FT	CONFLICT	499	499	K -> E (IN REF. 3).	
FT	CONFLICT	499	499	P -> G (IN REF. 2).	
FT	CONFLICT	527	527	D -> A (IN REF. 3).	
FT	CONFLICT	543	543	Q -> K (IN REF. 1).	
FT	CONFLICT	552	552	D -> V (IN REF. 1).	
FT	CONFLICT	643	643	A -> S (IN REF. 1).	
SQ	SEQUENCE	699 AA;	79258 MW;	ADD9697AE6AB0B5 CRC64;	
 Query Match 5.7% Score 151.5; DB 1; Length 699; Best Local Similarity 23.0%; Pred. No. 0.00054; Matches 48; Conservative 45; Mismatches 89; Indels 27; Gaps					
Qy	CGHVTSDSGTMTSKNYPGTYPTVCEKLTTPKCKRLRLRGD--NIETSK---TCAS 87				
Db	CSDNFRTGRVTISPPDPNPYPRKSSSECLYTILEEGPMVLAGEDFIDQHREVPCT 244				
Qy	185 DYLFFSSATDIQGYGCGSMANPKRELRLNSNYTLVLFKSGSHISRGFLTY--ASSHPD 145				
Db	245 DYIKRKGPKVLGFPGCGEKA--PEIISTOSHLLILFHSNDNAENRWRLSTRAGNCPE 303				
Qy	146 LITCL-----ERSHYEEKRYKFCPCAGCDIAGDISGNT-----KDGYRDLS-ICK 192				
Db	304 LQPHFGKIERSQAIFPKPDVLYSCDLGVKLVDNVEMDTFOIECLKDKGTWKNKITPC 363				
Qy	193 -----AIIHAGITTDDELGSINHLDLQSK 214				
Db	364 IYDCRAPGLEHGHLTFESTRNMLTTYSKE 392				
 RESULT 6 CLS_HUMAN STANDARD; PRT; 688 AA. CLS_HUMAN					
ID	CLS_HUMAN	STANDARD;	PRT;	688 AA.	
AC	P09871; Q9DMU4; Q9UCU7; Q9UCU8; Q9UCU9; Q9UCV0; Q9UCV1; Q9UCV2;				
AC	Q9UCV3; Q9UCV4; Q9UCV5;				
DT	01-MAR-1989 (Ref. 10, Created)				
DT	01-MAR-1989 (Ref. 10, Last sequence update)				
DE	15-JUN-2002 (Ref. 41, Last annotation update)				
GN	Complement C1s component precursor (EC 3.4.21.42) (C1 esterase).				
OS	C1s.				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=6901787; PubMed=2459702;				
RA	Kusunoto H., Hirotsawa S., Salter J.P., Hagen F.S., Kurachi K.;				
RT	"Human genes for complement components Clr and C1s in a close				
RL	tail-to-tail arrangement.";				
RP	Proc. Natl. Acad. Sci. U.S.A. 85:7307-7311(1988).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=88082788; PubMed=3500856;				
RA	McIntion C.M., Carter P.E., Smyth S.J., Dunbar B., Fothergill J.E.;				
RT	"Molecular cloning of cDNA for human complement component C1s. The				
RL	complete amino acid sequence.";				
RN	Eur. J. Biochem. 169:547-553(1987).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88163522; PubMed=2831944;				
RA	Tosti M., Duponchel C., Meo T., Jullier C.;				
RT	"Complete cDNA sequence of human complement C1s and close physical				
RL	linkage of the homologous genes C1s and C1r.";				
RN	Biochemistry 26:8516-8524(1987).				
RN	[4]				
RP	SEQUENCE OF 1-329 FROM N.A.				

RC TISSUE-Periphereal blood leukocytes;
 RA MEDLINE-9908558; PubMed-9794427;
 RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
 RA Nonaka M., Fujita T.;
 RT "Two lineages of mannose-binding lectin-associated serine protease
 RT (MASP) in vertebrates.";
 RT J. Immunol. 161:4924-4930(1998).
 RN [5]
 RP SEQUENCE OF 291-688 FROM N.A.
 RA MEDLINE-90040704; PubMed-2553984;
 RA Tosi M., Duponchel C., Meo T., Couture-Tosi E.;
 RT "Complement genes C1r and C1s feature an intronless serine protease
 RT domain closely related to haptoglobin.";
 RT J. Mol. Biol. 208:709-714(1989).
 RN [6]
 RP SEQUENCE OF 16-61, 168-219, 287-334 AND 384-445.
 RA MEDLINE-86164350; PubMed-3007145;
 RA Spycher S.E., Nick H., Rickli E.E.;
 RT "Human complement component C1s. Partial sequence determination of
 RT the heavy chain and identification of the peptide bond cleaved during
 RT activation.";
 RT Eur. J. Biochem. 156:49-57(1986).
 RN [7]
 RP SEQUENCE OF 438-500; 503-534; 542-601; 617-623 AND 626-656.
 RA MEDLINE-84104122; PubMed-6362661;
 RA Carter P.E., Dunbar B., Fothergill J.E.;
 RT "The serine protease chain of human complement component C1s.
 RT Cyanogen bromide cleavage and N-terminal sequences of the
 RT fragments.";
 RT Biochem. J. 215:565-571(1983).
 RN [8]
 RP PARTIAL SEQUENCE.
 RC TISSUE-Plasma;
 RA MEDLINE-91308095; PubMed-1854725;
 RA Ily C., Thieles N.M., Gagnon J., Arlaud G.J.;
 RT "Effect of lactoperoxidase-catalyzed iodination on the Ca(2+)-
 RT dependent interactions of human C1s. Location of the iodination
 RT sites.";
 RT Biochemistry 30:7135-7141(1991).
 RN [9]
 RP DISULFIDE BONDS.
 RA MEDLINE-91175725; PubMed-2007122;
 RA Hess D., Schaller J., Rickli E.E.;
 RT "Identification of the disulfide bonds of human complement C1s.";
 RT Biochemistry 30:2827-2833(1991).
 RN [10]
 RP PARTIAL SEQUENCE, AND 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
 RA MEDLINE-95398736; PubMed-7779774;
 RA Rossi V., Gaboriaud C., Lacroix M., Ulrich J., Fontecilla-Camps J.C.,
 RA Gagnon J., Arlaud G.J.;
 RT "Structure of the catalytic region of human complement protease C1s:
 RT study by chemical cross-linking and three-dimensional homology
 RT modeling.";
 RT Biochemistry 34:7311-7321(1995).
 RN [11]
 RP DISULFIDE.
 RA Dragon-Durey M.-A., Quartier P., Fremaux-Bacchi V., Biouin J.,
 RA de Barse C., Prieur A.-M., Weiss L., Fridman W.-H.;
 RT "Molecular basis of a selective C1s deficiency associated with early
 RT onset multiple autoimmune diseases.";
 RT J. Immunol. 166:7612-7616(2001).
 CC -1- FUNCTION: C1S B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1Q
 CC AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
 CC OF THE COMPLEMENT SYSTEM. C1R ACTIVATES C1S SO THAT IT CAN, IN
 CC TURN, ACTIVATE C2 AND C4.
 CC -1- CATALYTIC ACTIVITY: CLEAVES COMPONENT C4 TO C4A AND C4B, AND
 CC COMPONENT C2 TO C2A AND C2B.
 CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
 CC AND S IN THE MOLAR RATION OF 1:2:2. ACTIVATED C1S IS AN DISULFIDE-
 CC LINKED HETERODIMER OF AN HEAVY CHAIN AND A LIGHT CHAIN.
 CC -1- DISEASE: Defects in C1S are the cause of selective C1s deficiency,
 CC that is associated with early onset multiple autoimmune diseases.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: X06596; CAA29817.1; -
 DR EMBL: J04080; AAA51852.1; -
 DR EMBL: M18767; AAA51853.1; -
 DR EMBL: AB090767; BAA86864.1; -
 DR PIR: A40496; C1HDS.
 DR PIR: S00224; S00224.
 DR MEROPS: S01.193; -
 DR SWISS-2DPAGE: P09871; HDMAN.
 DR Genes: HGNC:1247; C1S.
 DR MIM: 120580; -
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00084; sushi_2.
 DR Pfam: PF00089; trypsin_1.
 DR Pfam: PF00431; CUB_2.
 DR PRINTS: PR00722; CHYOTRYPsin.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; TRYPSIN_DOM; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB_2; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; FALSE_NEG.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR Complement pathway; Plasma; Glycoprotein; Serine protease; Hydrolase;
 KW Hydroxylation; Sushi; Repeat; Signal; EGF-like domain;
 KW Calcium-binding.
 FT SIGNAL 1 15
 FT CHAIN 16 437
 FT CHAIN 438 688
 FT DOMAIN 16 130
 FT DOMAIN 131 172
 FT DOMAIN 175 290
 FT DOMAIN 293 355
 FT DOMAIN 358 422
 FT DOMAIN 438 688
 FT ACT_SITE 475 475
 FT ACT_SITE 529 529
 FT ACT_SITE 632 632
 FT MOD_RES 149 149
 FT CARBOHYD 174 174
 FT CARBOHYD 406 406
 FT DISULFID 65 83
 FT DISULFID 135 147
 FT DISULFID 143 156
 FT DISULFID 158 171
 FT DISULFID 175 202
 FT DISULFID 234 251
 FT DISULFID 341 341
 FT DISULFID 321 354

 CC COMPLEMENT C1S HEAVY CHAIN.
 CC CUB 1.
 CC EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 CC CUB 2.
 CC SUSHI 1.
 CC SUSHI 2.
 CC SERINE PROTEASE.
 CC CHARGE RELAY SYSTEM.
 CC CHARGE RELAY SYSTEM.
 CC HYDROXYLATION (PROBABLE).
 CC N-LINKED (GLCNAC. . .).
 CC N-LINKED (GLCNAC. . .).

```

FT DISULFID 359 403
FT DISULFID 386 421
FT DISULFID 425 549
FT DISULFID 595 618
FT DISULFID 628 659
FT CONFLICT 294 294
FT CONFLICT 513 513
FT CONFLICT 573 573
FT CONFLICT 645 646
SQ SEQUENCE 688 AA; 76684 MM; 85522647/44C47205 CRC64;

Query Match 5.7%; Score 151; DB 1: Length 688;
Best Local Similarity 25.3%; Pred. No. 0.00056;
Matches 46; Conservative 31; Mismatches 89; Indels 16; Gaps 5;

QY 15 FAVCAP--LRLQAEELDGGHVTSDSGTMTSPKPYNTYVEKIIIVKGRRL 71
DQ 155 FCSCEPREFLMDKNCNCVNGSDVFTALIGELINSPNPKYPERNSRCEYQIRLEKGRQV 214
QY 72 I--LRLGDLNIESKTCAS--DYLLFSSATDQYGYCG--SNAVPEKRLNNSNVTLPKRS 125
DQ 215 VVTLRREDFVEADSDAGNCNCLDSLVFAGDROFGPYCGHGPPLNITKSNALDIIFQT 274
QY 126 GSHISGRGFLTY-----ASSDHPDLITCLERSHFEERYSKFCFAGCDIAGDISG 178
DQ 275 DLFGQKGMKRLRHGDPMPCPEKREDTPNSWEPAPKAKYFRDVQITCLDGEVEGRVGA 334
QY 179 NT 180
DQ 335 TS 336

RESULT 7
CIR_HUMAN STANDARD; PRT; 705 AA.
ID CIR_HUMAN
AC P00736;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement C1r component precursor (EC 3.4.21.41).
GN CIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE-87026566; PubMed-3021205;
RA Leytus S.P., Kurachi K., Sakariassen K.S., Davie E.W.;
RL "Nucleotide sequence of the cDNA coding for human complement C1r.";
Biochemistry 25:4855-4863(1986).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE-87156625; PubMed-3030286;
RA Journet A., Tosi M.;
RL "Cloning and sequencing of full-length cDNA encoding the precursor of
human complement component C1r.";
Biochem. J. 240:783-787(1986).
[3]
RP SEQUENCE OF 18-463.
RX MEDLINE-87241248; PubMed-3036070;
RA Arlaud G.J., Willis A.C., Gagnon J.;
RL "Complete amino acid sequence of the A chain of human complement-
classical-pathway enzyme C1r.";
Biochem. J. 241:711-720(1987).
[4]
RP SEQUENCE OF 464-705.
RX MEDLINE-83204782; PubMed-6303394;
RA Arlaud G.J., Gagnon J.;
RL "Complete amino acid sequence of the catalytic chain of human
complement subcomponent C1-r.";
Biochemistry 22:1758-1764(1983).
[5]

```

```

RP SEQUENCE OF 152-186, AND HYDROXYLATION.
RX MEDLINE-88005128; PubMed-2820791;
RA Arlaud G.J., van Dorsselaer A., Bell A., Mancini M., Aude C.,
RA Gagnon J.;
RL "Identification of erythro-beta-hydroxyasparagine in the EGF-like
domain of human C1r.";
FEBS Lett. 222:129-134(1987).
[6]
RP SEQUENCE OF 133-137; 187-211 AND 609-613, AND PHOSPHORYLATION.
RX MEDLINE-96221263; PubMed-8635594;
RA Pelloux S., Thielemans N.M., Hudry-Clergeon G., Petillot Y., Filhol O.,
RA Arlaud G.J.;
RL "Identification of a cryptic protein kinase CK2 phosphorylation site
in human complement protease C1r, and its use to probe intramolecular
interaction.";
FEBS Lett. 386:15-20(1996).
[7]
RP STRUCTURE BY NMR OF 140-192.
RX MEDLINE-98138432; PubMed-9477945;
RA Bersch B., Hernandez J.-F., Marion D., Arlaud G.J.;
RL "Solution structure of the epidermal growth factor (EGF)-like module
of human complement protease C1r, an atypical member of the EGF
family.";
Biochemistry 37:1204-1214(1998).
CC -1- FUNCTION: C1R B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1Q
AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
OF THE COMPLEMENT SYSTEM.
CC -1- CATALYTIC ACTIVITY: ACTIVATES C1S BY PROTEOLYTIC CLEAVAGE SO
THAT IT CAN, IN TURN, ACTIVATE C2 AND C4.
CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
AND S IN THE MOLAR RATION OF 1:2:2. C1R IS A DIMER OF IDENTICAL
CHAINS, EACH OF WHICH IS ACTIVATED BY CLEAVAGE INTO TWO CHAINS, A
AND B, CONNECTED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04701; CAA28407.1; -
DR PIR; A24170; C1HHRB
DR PDB; 1APQ; 17-SEP-97.
DR MEROPS; S01.192; -.
DR GeneW; HGNC:1246; C1R.
DR MIM; 216950; -
DR InterPro; IPR000152; Asx_hydroxyl-
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 2.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00431; CUB; 2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.

```

DR PROSITE; PS50240; TRYPsin.DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Complement pathway: Plasma: Glycoprotein: Serine protease; Hydrolase;
 KW Hydroxylation: Phosphorylation; Sush1: Repeat; Signal;
 KW EGF-like domain; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 463 COMPLEMENT C1R HEAVY CHAIN.
 FT CHAIN 18 464 COMPLEMENT C1R LIGHT CHAIN.
 FT DOMAIN 142 190 CUB 1.
 FT DOMAIN 142 190 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 193 305 CUB 2.
 FT DOMAIN 308 372 SUSH1 1.
 FT DOMAIN 375 448 SUSH1 2.
 FT DOMAIN 464 705 SERINE PROTEASE.
 FT ACT SITE 502 502 CHARGE RELAY SYSTEM.
 FT ACT SITE 557 557 CHARGE RELAY SYSTEM.
 FT ACT SITE 654 654 CHARGE RELAY SYSTEM.
 FT MOD_RES 167 167 HYDROXYLATION.
 FT MOD_RES 206 206 PHOSPHORYLATION (BY CK2).
 FT CARBOHD 125 125 N-LINKED (GLCNAC. . .).
 FT CARBOHD 221 221 N-LINKED (GLCNAC. . .).
 FT CARBOHD 514 514 N-LINKED (GLCNAC. . .).
 FT CARBOHD 581 581 N-LINKED (GLCNAC. . .).
 FT DISULFD 71 89 PROBABLE.
 FT DISULFD 146 165
 FT DISULFD 161 174
 FT DISULFD 176 189
 FT DISULFD 193 220
 FT DISULFD 250 268 PROBABLE.
 FT DISULFD 309 358 PROBABLE.
 FT DISULFD 338 371 PROBABLE.
 FT DISULFD 376 429 PROBABLE.
 FT DISULFD 406 447 PROBABLE.
 FT DISULFD 451 577 INTERCHAIN (PROBABLE).
 FT DISULFD 620 639 POTENTIAL.
 FT DISULFD 650 680
 FT CONFLICT 152 152
 FT SEQUENCE 705 AA; 80173 MW; 5CBCC0201061463 CRC64;
 Query Match 5.5%; Score 147; DB 1; Length 705;
 Best Local Similarity 27.7%; Pred. No. 0.0012;
 Matches 52; Conservative 30; Mismatches 84; Indels 22; Gaps 7;

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cochlea;
 RX MEDLINE=98110569; PubMed=9441737;
 RA Robertson N.G., Skvorak A.B., Yin Y., Weremowicz S., Johnson K.R.,
 RA Kovach K.A., Batley J.E., Bleher F.R., Morton C.C.;
 RA "Mapping and characterization of a novel cochlear gene in human and in
 mouse: a positional candidate gene for a deafness disorder, DFNA9.",
 RL Genomics 46:345-354(1997).
 RN [2]
 RP VARIANTS DFNA9 GLY-66; GLU-88 AND ARG-117.
 RX MEDLINE=99021390; PubMed=9806553;
 RA Robertson N.G., Lu L., Heller S., Merchant S.N., Bavey R.D.,
 RA McKenna M., Nadol J.B. Jr., Miyamoto R.T., Linthicum F.H. Jr.,
 RA Neto J.F.L., Hudspeth A.J., Seidman C.E., Morton C.C., Seidman J.G.;
 RA "Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic
 deafness with vestibular dysfunction.",
 RL Nat. Genet. 20:299-303(1998).
 RN [3]
 RP VARIANT DFNA9 SER-51.
 RX MEDLINE=99135917; PubMed=991344;
 RA de Kok Y.J.M., Bom S.J.H., Brunt T.M., Kemperman M.H.,
 RA van Beusekom E., van der Velde-Visser S.D., Robertson N.G.,
 RA Morton C.C., Huygen P.L.M., Verhagen W.I.M., Brunner H.G.,
 RA Cremers J.W.R.J., Cremers F.P.M.;
 RA "A proxi-to-ser mutation in the COCH gene is associated with late
 onset autosomal dominant progressive sensorineural hearing loss with
 vestibular defects.",
 RL Hum. Mol. Genet. 8:361-366(1999).
 RN [4]
 RP VARIANT DFNA9 ASN-109.
 RA Kamariotis M., McGill J., Lynch M., Dahl H.-H.M.;
 RA "Genetic studies of a family with inherited dominant hearing loss.",
 RL Unpublished observations (JAN-2000).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN INNER EAR STRUCTURES; THE COCHLEA
 AND THE VESTIBULE.
 CC -1- DISEASE: DEFECTS IN COCH ARE THE CAUSE OF AUTOSOMAL DOMINANT
 NONSYNDROMIC SENSORINEURAL DEAFNESS TYPE 9 (DFNA9). AFFECTED
 INDIVIDUALS HAVE MICROPOLYSACCHARIDE DEPOSITIONS IN THE CHANNELS OF
 THE COCHLEAR AND VESTIBULAR NERVES. THESE DEPOSITIONS APPARENTLY
 CAUSE STRANGULATION AND DEGENERATION OF DENDRITIC FIBERS.
 CC -1- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
 CC -1- DATABASE: NAME-Protein Spotlight;
 NOTE-Issue 4 of November 2000;
 WWW="http://www.expasy.org/spotlight/articles/spilt004.html";
 CC -1- DATABASE: NAME-Hereditary hearing loss homepage;
 WWW="http://www.uia.ac.be/dna1ab/hhh/hhgenes.html";
 CC NOTE-Genes page;
 CC WWW="http://www.uia.ac.be/dna1ab/hhh/hhgenes.html";
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb.ch).
 CC -----
 CC EMBL; AF006740; AAC39545.1; -
 DR HSSP; P11215; LJLM.
 DR Genew; HGNC:2180; COCH.
 DR MIM; 603196; -
 DR MIM; 601369; -
 DR InterPro; IPR004043; LCCL_dom.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00092; vwa; 2.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00327; VWA; 2.
 DR PROSITE; PS50820; LCCL; 1.
 DR PROSITE; PS50334; VWFA; 2.
 KW Glycoprotein; Repeat; Signal; Disease mutation; Deafness;


```

OY 187 TSLCKAAIHAGITT--DELGHINILQSK-----GISHYE-----GLT-----223
DB 286 -----IHSDQITASSQGTNNVSERSLANTPENCGTGEDSGKYMIOVDGLLFEVT 337
OY 224 ---ANGVLSHSGLSERKLEFTPGAMITVALPSYF-----TALLITG-----MGI 268
DB 338 AVGTGQAISE---TRKKYVVKTYRVDISSNGEDWISLKEGNAKIIFOGWTNPDLVLCV 394
OY 269 FA-----ICRRKKKGNPNVSADAKTGCKOIKYFPARH-----OSTETISYDN 314
DB 395 FSKPLTRFPRIRKIPVSENGISMRFEYGC-KITDIPCSGLMGLVSLIDSOITASSNOA 453
OY 315 EKE-WTOKLDLITS-----DNADVO--OPLMITGS-----341
DB 454 DRMMENIRLTSRTGMALPSPHPYTNEMLQVDLDEKIVAGVILQGGKHENKVEK 513
OY 342 ---TAAKSGTFRPMDTOEEVRVNTASGHDCRRPRHREXALPLTISEPEYATPT- 396
DB 514 KFIAYSNMGSDMKTTIMDSKRRAKSFEGNMNDTPELRTFSPLSTRFIRIYERATHSG 573
OY 397 --VERHLRAHFTSGSYVPGPRPTHK-----HSHSGGFPPATGATOVESYORP 446
DB 574 LGIRMLLGEVAPTAGPTTPNGNVHECDDOANCHSGTGDF-QLTGCTVLAIEKP 632

RESULT 10
BMPL_MOUSE
ID BMPL_MOUSE STANDARD: PRT: 991 AA.
AC P98063:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bone morphogenetic protein 1 precursor (EC 3.4.24.19) (BMP-1)
GN (Procollagen C-proteinase) (PCP) (Mammalian tolloid protein) (mtld).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RZ MEDLINE=94229342; PubMed=8174772;
RA Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;
RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),
RT which is related to the Drosophila dorsventral gene tolloid and
RT encodes a putative astacin metalloendopeptidase."
RL Dev. Biol. 163:175-183(1994).
CC -1- FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II
CC AND II. INDUCES CARTILAGE AND BONE FORMATION.
CC -1- CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at
CC Ala-I-Asp in type I and II procollagens and at Arg-I-Asp in type
CC III.
CC -1- CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-
CC ENDOPEPTIDASE ENHANCER PROTEIN.
CC -1- TISSUE SPECIFICITY: AT HIGH LEVELS IN EMBRYONIC MATERNAL DECIDUUM
CC AND FLOOR PLATE REGION OF THE NEURAL TUBE. LESS IN DEVELOPING
CC MEMBRANOUS AND ENDOCHONDRAL BONE, SUBMUCOSA OF INTESTINE, DERMIS
CC OF SKIN AND THE MESENCHYME OF SPLEEN AND LUNG.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
CC -1- SIMILARITY: CONTAINS 5 CUB DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L24755; AAA37306.1; -

```

```

DR HSSP; P00736; IAPQ.
DR MEROPS; M12.005; -.
DR MGD; MGI:88176; Bmpl1.
DR InterPro; IPR001506; Astacin.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00431; CUB_5.
DR Pfam; PF01400; Astacin; 1.
DR PRINTS; PR00480; ASTACIN.
DR SMART; SM00042; CUB_5.
DR SMART; SM00179; EGF_CA_2.
DR SMART; SM00235; ZMGC_1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00180; CUB_5.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01187; EGF_CA_2.
DR Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrilase;
KW Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;
KW Glycoprotein; Zymogen.
FT SIGNAL; 1 25
FT PROPEP; 26 125
FT CHAIN; 126 991
FT DOMAIN; 126 326
FT DOMAIN; 327 439
FT DOMAIN; 440 551
FT DOMAIN; 552 593
FT DOMAIN; 596 707
FT DOMAIN; 708 748
FT DOMAIN; 752 864
FT DOMAIN; 865 981
FT METAL; 218 218
FT ACT_SITE; 219 219
FT METAL; 222 222
FT METAL; 228 228
FT DISULFID; 327 333
FT DISULFID; 380 402
FT DISULFID; 440 466
FT DISULFID; 493 515
FT DISULFID; 556 568
FT DISULFID; 564 577
FT DISULFID; 579 592
FT DISULFID; 596 622
FT DISULFID; 649 671
FT DISULFID; 712 723
FT DISULFID; 719 732
FT DISULFID; 734 747
FT CARBOHYD; 96 96
FT CARBOHYD; 147 147
FT CARBOHYD; 337 337
FT CARBOHYD; 368 368
FT CARBOHYD; 604 604
SQ SEQUENCE 991 AA; 111607 MW; 68A1847783A0BB9E CRC64;

```

```

Query Match 5.3%; Score 142.5; DB 1; Length 991;
Best Local Similarity 33.0%; Pred. No. 0.0044;
Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;

```

```

OY 31 GCCHITSDSGMTSKNTPGTYPTVCERKITVKGKRLILRLDNLNISK-TASD 89
DB 751 GCEHKYTS-TSGHITSPNWDKPSKRECKWASSTPGHVRVKTLPVEMDIESOPBCAYDH 809
OY 90 LIFSSATDQGP-----YCGSWAVPKELRLNSNVTVLFSGSHISGRGFLTYAS 140
DB 810 LEVFDGRDAKAPVLRGFCGS-KRPEPVLATGNMFLRFTSDNSVQKKGQASHST 863

```

RESULT 11

FT VARSPLIC 589 622 AACGGTUKINGSITSPGKPEPPNKNCTMOLV -> GCY
 FT DLOVGRPLMDRHCFLSTGPEMLGTALNG (IN
 FT ISOFORM BMP1-5).
 FT VARSPLIC 623 986 MISSING (IN ISOFORM BMP1-5).
 FT VARSPLIC 703 717 DKDECSKNGGCCOD -> GGELFGILGHPRRP (IN
 FT ISOFORM BMP1-6).
 FT VARSPLIC 718 986 MISSING (IN ISOFORM BMP1-6).
 FT VARSPLIC 703 823 DKDECSKNGGCCODVNTGSECCQSGFVLHKNKDK
 FT EACGDHVTSTGCTISPMMPDKYPSKRECTMAISYPIGR
 FT VKLTFEMDIESQPECAVDHLEVDGDAAAPVLRGCG
 FT -> VLKAGDRHSHLSGLLELPHALVDVPPASALHD
 FT THAHTHTHTCPIAOTRCRGPPGLASRLSPGRLTIA
 FT POGSTLDFDTIRGDPKPRKRKSLTSTPATFERGIVA
 FT L (IN ISOFORM BMP1-7).
 FT VARSPLIC 824 986 MISSING (IN ISOFORM BMP1-7).
 FT CONFLICT 748 748 D -> N (IN REF. 4).
 FT CONFLICT 934 934 R -> S (IN REF. 4).
 SO SEQUENCE 986 AA; 111248 MW; F89201913AC3CEBA CRC64;

Query Match 5.3%; Score 141.5; DB 1; Length 986;
 Best Local Similarity 33.0%; Pred. No. 0.0052;
 Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;

QY 31 CGGHIVTSQDSGWTSTKPYPTVYCEKIIIVPKGRILRLGLDNLIESK-TGASDY 89
 DB 746 GCHKRTS-TSGITSPMDKPKSKCEKTAISSTPGHRYKLFEMEDIESQPECAVDH 804
 QY 90 LFFSATDQYGP---YCGSMAVPEKELRLNSNEVTVFKSGSHISGSGFLITYVS 140
 DB 805 LEVFDGDAAAPVLRGCGS-KRPEVPLATGSRKFLRFYSDNSVQRGQASHTAT 858

RESULT 12
 UVS2_XENLA STANDARD; PRT; 514 AA.
 AC P42664;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Embryonic protein UVS.2 precursor (EC 3.4.24.-).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97228908; PubMed=9074934;
 RA Katagiri C., Maeda R., Yamashita C., Mita K., Sargent T.D.,
 RA Yasunasu S.;
 RT "Molecular cloning of Xenopus hatching enzyme and its specific
 RT expression in hatching gland cells.";
 RT Int. J. Dev. Biol. 41:19-25(1997).
 RN [2]
 RP SEQUENCE OF 196-514 FROM N.A.
 RX MEDLINE=90108413; PubMed=1688538;
 RA Sato S.M., Sargent T.D.;
 RT "Molecular approach to dorsoanterior development in Xenopus laevis.";
 RT Dev. Biol. 137:135-141(1990).
 RL [1]
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: EXCLUSIVELY IN THE ANTERIOR NEURAL FOLD OF
 CC NEURULA STAGE EMBRYOS. BY THE TAILBUD STAGE, THE PROTEIN IS
 CC LOCALIZED IN SPECIALIZED CEPHALIC ECTODERM, IN A REGION PROBABLY
 CC CORRESPONDING TO THE HATCHING GLAND.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; D89632; BA014003.1; -
 CC DR EMBL; M27162; AAA49980.1; -
 CC DR HSSP; P07584; 11AE.
 CC DR MEROPS; M12.014; -
 CC DR Interpro; IPR001506; Astacin.
 CC DR Interpro; IPR000859; CUB_domain.
 CC DR Interpro; IPR000130; Zn_MTPetldse.
 CC DR Pfam; PF00431; CUB; 2.
 CC DR Pfam; PF01400; Astacin; 1.
 CC DR PRINTS; PR00480; ASTACIN.
 CC DR SMART; SM00042; CUB; 2.
 CC DR SMART; SM00235; ZMNC; 1.
 CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC DR PROSITE; PS01180; CUB; 2.
 CC KW Hydroxylase; Metalloprotease; Zinc; Repeat; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 514
 FT DOMAIN 288 400 EMBRYONIC PROTEIN UVS.2.
 FT DOMAIN 402 513 CUB 1.
 FT METAL 186 186 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 187 187 BY SIMILARITY.
 FT METAL 190 190 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 196 196 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 514 AA; 56844 MW; DE1FCB599014D986 CRC64;

Query Match 5.3%; Score 141; DB 1; Length 514;
 Best Local Similarity 34.5%; Pred. No. 0.0023;
 Matches 38; Conservative 11; Mismatches 57; Indels 4; Gaps 3;

QY 32 CGGHIVTSQDSGWTSTKPYPTVYCEKIIIVPKGRILRLGLDNLIE-SKTCASDYL 90
 DB 402 CGCAFYSPPK-TFSPVNPVNTTNCWTITAPAGFKVSLRTDELEIGASCARDYL 460
 QY 91 LFFSAT--DOYPCGSMVPEKELRLNSNEVTVFKSGSHISGSGFLITY 138
 DB 461 NITVSTLGAWPGYCGPIDFHSALIVSKNSMTMTNSDFSKOYKGFSAATY 510

RESULT 13
 BMPH_STRPU STANDARD; PRT; 639 AA.
 AC P98069;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bone morphogenetic protein 1 homolog precursor (EC 3.4.24.-) (SUBMP).
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinozoa; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94215496; PubMed=8162855;
 RA Hwang S.P.L., Parin J.S., Lennarz W.J.;
 RT "Characterization of a homolog of human bone morphogenetic protein 1
 RT in the embryo of the sea urchin, Strongylocentrotus purpuratus.";
 RT Development 120:559-568(1994).
 RL [1]
 CC -1- TISSUE SPECIFICITY: ECTODERMAL AND PRIMARY MESENCHYM CELLS IN
 CC HATCHED BLASTULA.
 CC -1- DEVELOPMENTAL STAGE: EMBRYO, HIGHEST LEVEL BEFORE SPICULOGENESIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; L23838; AAA30081.1; -
 DR HSSP; P00736; IAPQ.
 DR MEROPS; M12.005; -
 DR InterPro; IPR001506; Astacin.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_ca.
 DR InterPro; IPR000130; Zn_MTPeptide.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF01400; Astacin; 1.
 DR PRINTS; PR00480; ASTACIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00235; ZNMG; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 KW Developmental protein; Repeat; Hydrolase; Protease; Zinc; Embryo;
 KW Metalloprotease; EGF-like domain; Calcium; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 109
 FT CHAIN 110 639
 FT DOMAIN 110 306
 FT DOMAIN 307 419
 FT DOMAIN 420 531
 FT DOMAIN 532 573
 FT METAL 197 197
 FT ACT_SITE 198 198
 FT METAL 201 201
 FT METAL 207 207
 FT DISULFID 536 548
 FT DISULFID 544 557
 FT DISULFID 559 572
 FT CARBOHYD 122 122
 FT CARBOHYD 140 140
 FT CARBOHYD 317 317
 FT CARBOHYD 455 455
 SEQUENCE 639 AA; 71893 MW; 59307B265B7894AD CRC64;
 Query Match 5.3%; Score 141; DB 1; Length 639;
 Best Local Similarity 26.2%; Pred. No. 0.0031;
 Matches 50; Conservative 26; Mismatches 63; Indels 50; Gaps 10;

RESULT 14
 ID PCOL_RAT STANDARD; PRT; 468 AA.
 AC 008628;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I
 DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-
 DE proteinase enhancer protein).
 GN PCOLCE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; PubMed=9303490;
 RX MEDLINE=97447750; PubMed=9303490;
 RA Ogata I., Auster A.S., Matsui A., Greenwel P., Geerts A., D'Amico T.,
 RA Fujiwara K., Kessler E., Rojkind M.;
 RT "Up-regulation of type I procollagen C-proteinase enhancer protein
 RT messenger RNA in rats with CCl4-induced liver fibrosis.";
 RL Hepatology 26:611-617(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Aorta;
 RA Hirahara I., Syoufuda K., Harada K., Tomita M., Urakami K., Terai H.,
 RA Morisaki N., Saito Y.;
 RT "Smooth muscle cell derived procollagen C-protease enhancer protein.";
 RL Cell Struct. Funct. 21:662-662(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344;
 RA Masuda M., Igarashi H., Kano M., Yoshikura H.;
 RT "Proximal integration into the procollagen C-proteinase enhancer
 RT protein gene and its effects in cultured rat fibroblasts revealed by
 RT an excisable 'hit-and-run' retroviral vector.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=91076903; PubMed=2256940;
 RA Kessler E., Mould A.P., Holmes D.J.S.;
 RT "Procollagen type I C-proteinase enhancer is a naturally occurring
 RT connective tissue glycoprotein.";
 RL Biochem. Biophys. Res. Commun. 173:81-86(1990).
 CC -1- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
 CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN COLLAGEN-RICH
 CC TISSUES, ESPECIALLY TENDON. ALSO EXPRESSED IN CORNEA AND STERNA.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

FT DISULFID 765 778 BY SIMILARITY.
FT DISULFID 783 809 BY SIMILARITY.
FT DISULFID 836 858 BY SIMILARITY.
FT DISULFID 896 926 BY SIMILARITY.
FT DISULFID 953 975 BY SIMILARITY.
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1022 AA; 115536 MW; A6BCALDDE41793F9 CRC64;

Query Match 5.28; Score 139; DB 1; Length 1022;
Best Local Similarity 29.08; Pred. No. 0.0086;

Matches 40; Conservative 23; Mismatches 45; Indels 30; Gaps 7;

QY 32 CGHIVTSODS-GTMTSKNYPGYPNYTVCEKITVPRKRLILRLGLDNI-ESKTCASDY 89
DB 358 CGE--TLDDSVGNFSSPGTIPNGIPSTHCGWRIISVTPGKIVLNTFTMDLYKSSLCWYDY 415
QY 90 LFFSSATDQY-----GPGCGMAVPKELRLNSNEVTVLFRKSGSHISGRFLTY---- 138
DB 416 I---EVROGYNRKAPLGRFGD-KIPEYLVSTDSNMWIEFRSSSNWVGCPFAAYEAIC 471
QY 139 -----ASSDHPD 145
DB 472 GGEISKDSGOIOSPNYPD 489

Search completed: May 15, 2003, 13:16:10
Job time : 17.6281 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 49.1437 Seconds
(without alignments)
2108.953 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPSVIALFLFVCAAP.....YSAPRNGIAPLNGTAMTALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPRMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2665	99.8	503	11	Q9DAJ3 mus musculu
2	2448	91.7	460	11	Q9D696 mus musculu
3	2239.5	83.8	432	11	Q9B327 mus musculu
4	575.5	21.5	775	4	Q96PD2 mus sapien
5	575.5	21.5	775	4	Q8TDX2 mus sapien
6	565.5	21.2	769	11	Q91ZV2 rattus norv
7	560	21.0	769	11	Q91ZV3 mus musculu
8	503	18.8	155	4	Q96NH2 mus sapien
9	418	15.6	364	4	Q14089 homo sapien
10	178	6.7	414	11	Q9CX06 mus musculu
11	178	6.7	414	11	Q9CX06 mus musculu
12	172	6.4	414	11	Q8BY4 mus musculu
13	170.5	6.4	415	4	Q9UKZ9 mus musculu
14	169.5	6.3	694	11	Q70542 rattus norv
15	169.5	6.3	694	11	Q8R099 mus musculu
16	168	6.3	1524	13	Q91674 xenopus lae

17	167.5	6.3	3623	11	Q70244 rattus norv
18	165	6.2	541	11	Q9GX90 rattus norv
19	165	6.2	623	11	Q9JJP3 rattus norv
20	165	6.2	643	11	Q9GX84 rattus norv
21	165	6.2	678	11	Q9JUS8 rattus norv
22	162.5	6.1	3623	4	Q60494 mus sapien
23	162	6.1	685	11	Q92338 mus musculu
24	160	6.0	685	11	Q91WPO mus musculu
25	159.5	6.0	701	11	Q9JUS9 rattus norv
26	159.5	6.0	733	11	Q92050 mus musculu
27	159.5	6.0	746	5	Q01654 halocynthia
28	157.5	5.9	3620	6	Q9T053 canis famill
29	157	5.9	276	4	Q9BRH3 cyprinus ca
30	157	5.9	384	13	Q9W630 cyprinus ca
31	154	5.8	251	11	Q9D9K5 mus musculu
32	154	5.8	745	13	Q9PVY3 cyprinus ca
33	153.5	5.7	719	13	Q9PVY2 triakis scy
34	152.5	5.7	919	13	Q8UVR0 gallus gall
35	151.5	5.7	936	13	Q8UVQ9 oryctolagus
36	151.5	5.7	1594	6	Q95218 homo sapien
37	150	5.6	926	4	Q9UQ00 homo sapien
38	148	5.5	1015	4	Q9Y6L7 homo sapien
39	147	5.5	494	4	Q96DM8 homo sapien
40	145	5.4	693	4	Q96DM8 homo sapien
41	145	5.4	722	13	Q9PSZ5 lampetra ja
42	144.5	5.4	650	11	Q8VHT5 mus musculu
43	144	5.4	977	13	Q91925 xenopus lae
44	144	5.4	579	4	Q96DQ9 homo sapien
45	143	5.4			

ALIGNMENTS

RESULT 1
ID Q9DAJ3 PRELIMINARY; PRT; 503 AA.
AC Q9DAJ3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 4631413K1R1K protein.
GN 4631413K1R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batdalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehle H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hashizaki Y.;
Functional annotation of a full-length mouse cDNA collection.;
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK016485; BAB30265.1; -.

DR MGD: MGI:1913936; 4631413K1R1K.
 DR InterPro: IPR000859; CUB_dom.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SO SEQUENCE 503 AA; 54547 MW; F8B121E845CA06B CRC64;

Query Match 99.8%; Score 2665; DB 11; Length 503;
 Best Local Similarity 99.8%; Pred. No. 1.5e-219;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MGTAGGSPVALAFVACAPLRQAEEIGDGGHVTSDSGTMSKYPCTPYATYCE 60
 1 MGTAGGSPVALAFVACAPLRQAEEIGDGGHVTSDSGTMSKYPCTPYATYCE 60
 61 KITVPKGRLLRLRGLDNIESKTCASDYLLFSSATDQYCGSWAVPKELRLNSNEVT 120
 61 KITVPKGRLLRLRGLDNIESKTCASDYLLFSSATDQYCGSWAVPKELRLNSNEVT 120
 121 VLFSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSKFCPCACRDIAIGISGNT 180
 121 VLFSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSKFCPCACRDIAIGISGNT 180
 181 KDGRDPSLCKAIIHAGIITDELGHINILQSGISHYEGILANGVLSRRGISEKREFL 240
 181 KDGRDPSLCKAIIHAGIITDELGHINILQSGISHYEGILANGVLSRRGISEKREFL 240
 241 FTFPGMNTTVAIPSVIFALLLTGMGIFALCRKRKKGNPYVSDAQKTCWKQIKYKF 300
 241 FTFPGMNTTVAIPSVIFALLLTGMGIFALCRKRKKGNPYVSDAQKTCWKQIKYKF 300
 301 ARHSTETITSDNEKEKTKQLDITSDMADYQOPLMIGTGVARKSGTFPMOTDEEV 360
 301 ARHSTETITSDNEKEKTKQLDITSDMADYQOPLMIGTGVARKSGTFPMOTDEEV 360
 361 RVNTEASGHYCCPRPHRHEVALPLTSEPEYATPIVERHLLRAHTSTOSGYRPPGRP 420
 361 RVNTEASGHYCCPRPHRHEVALPLTSEPEYATPIVERHLLRAHTSTOSGYRPPGRP 420
 421 THKSHSSGGFPATGATQVESYORPASPKPVGGYDPAASSFLDSNDPASQSMISGG 480
 421 THKSHSSGGFPATGATQVESYORPASPKPVGGYDPAASSFLDSNDPASQSMISGG 480
 481 DDGYSAPRNGIAPLNGTAMTALL 503
 481 DDGYSAPRNGIAPLNGTAMTALL 503

RESULT 2
 ID 09D696 PRELIMINARY; PRT; 460 AA.
 AC 09D696;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 4631413K1R1K protein.
 GN 4631413K1R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shlnagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Hono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 FT *Functional annotation of a full-length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK014521; BAB29409.1;
 DR MGD: MGI:1913936; 4631413K1R1K.
 DR InterPro: IPR000859; CUB_dom.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SO SEQUENCE 460 AA; 50334 MW; 3CF3356F70A88B93 CRC64;

Query Match 91.7%; Score 2440; DB 11; Length 460;
 Best Local Similarity 100.0%; Pred. No. 5.1e-201;
 Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

44 MTSKNYPGTNYNYCEKITVPKGRLLRLRGLDNIESKTCASDYLLFSSATDQYGYPC 103
 1 MTSKNYPGTNYNYCEKITVPKGRLLRLRGLDNIESKTCASDYLLFSSATDQYGYPC 60
 104 GSMVAPKELRLNSNEVTYLFKSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSK 163
 61 GSMVAPKELRLNSNEVTYLFKSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSK 120
 164 FCPAGCRDIADISGNKRDGRDPSLCKAIIHAGIITDELGHINILQSGISHYEGIL 223
 121 FCPAGCRDIADISGNKRDGRDPSLCKAIIHAGIITDELGHINILQSGISHYEGIL 180
 224 ANGVLSRRGSLSEKREFLTPGNNITVAIPSVIFALLLTGMGIFALCRKRKKGNPYV 283
 181 ANGVLSRRGSLSEKREFLTPGNNITVAIPSVIFALLLTGMGIFALCRKRKKGNPYV 240
 284 SADAKTCWKQIKYPPARHSTETITSDNEKEKTKQLDITSDMADYQOPLMIGTGV 343
 241 SADAKTCWKQIKYPPARHSTETITSDNEKEKTKQLDITSDMADYQOPLMIGTGV 300
 344 ARKSTFPMOTDEEVYRVNTEASGHYDCPRPHRHEVALPLTSEPEYATPIVERHLLR 403
 301 ARKSTFPMOTDEEVYRVNTEASGHYDCPRPHRHEVALPLTSEPEYATPIVERHLLR 360
 404 AHPTSTOSGYRVPPRPLTHKSHSSGGFPATGATQVESYORPASPKPVGGYDPAASS 463
 361 AHPTSTOSGYRVPPRPLTHKSHSSGGFPATGATQVESYORPASPKPVGGYDPAASS 420
 464 FLDSRDPASQSMISGGDDGYSAPRNGIAPLNGTAMTALL 503
 421 FLDSRDPASQSMISGGDDGYSAPRNGIAPLNGTAMTALL 460

RESULT 3
 ID 08R327 PRELIMINARY; PRT; 432 AA.
 AC 08R327;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to Riken cDNA 4631413K11 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC026771; AAH26771.1;
 SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 83.8%; Score 2239.5; DB 11; Length 432;
 Best Local Similarity 85.5%; Pred. No. 3,3e-183;
 Matches 430; Conservative 1; Mismatches 1; Indels 71; Gaps 1;

QY 1 MGAGGAPSYLALFANCAPLRLOAEELGDCGHIYVSODSGITMSKPNPTNYNYVCE 60
 Db 1 MGAGGAPSYLALFANCAPLRLOAEEL----- 28
 QY 61 KIIVPAGKRLILRLGDLNIESKTCASDYLLFSATDOYGPYCGSMAPKELRLNSMEY 120
 Db 29 -----GPGCGSMAPKELRLNSMEY 49
 QY 121 VLEKSGSHISGRGFLTYASDHPDLITCLERGSHYFEERYSKFCPACGCDIADISGNT 180
 Db 50 VLEKSGSHISGRGFLTYASDHPDLITCLERGSHYFEERYSKFCPACGCDIADISGNT 109
 QY 181 KQYRQSLCKAIIHAGIITDELGHIINLQSKGISHYEGLLANGVLSRHSLSERKFL 240
 Db 110 KQGRQSLCKAIIHAGIITDELGHIINLQSKGISHYEGLLANGVLSRHSLSERKFL 169
 QY 241 FTTPGNNITVAIPSVYFIALLLTGMIFAICRRKKRKNPYVSADAKTGCKQIYYPF 300
 Db 170 FTTPGNNITVAIPSVYFIALLLTGMIFAICRRKKRKNPYVSADAKTGCKQIYYPF 229
 QY 301 ARHOSFTTISYNEKEMTKDLITSDMADYQPLMIGTVARKSGTFRPMDTIEEV 360
 Db 230 ARHOSFTTISYNEKEMTKDLITSDMADYQPLMIGTVARKSGTFRPMDTIEEV 289
 QY 361 RVNTEASGHDCPRPGRHEVALPLTSHSEPEYAPIYERHLRLNHTSGTSGYVGPBP 420
 Db 290 RVNTEASGHDCPRPGRHEVALPLTSHSEPEYAPIYERHLRLNHTSGTSGYVGPBP 349
 QY 421 THKSHSGGFPATGATQVESYORPASPPVGGYDKPASSELDRDASQSOMTSGG 480
 Db 350 THKSHSGGFPATGATQVESYORPASPPVGGYDKPASSELDRDASQSOMTSGG 409
 QY 481 DDCYSAPRNLAPLNGATALL 503
 Db 410 DDCYSAPRNLAPLNGATALL 432

RESULT 4
 Q96PD2 PRELIMINARY; PRT; 775 AA.
 AC Q96PD2;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 GN ESDN.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX PubMed-11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the Longest Secretory Signal Sequence among Eukaryotes, is Up-regulated after Vascular Injury.";
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1. SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 DR EMBL: AF387547; AAL30178.1; -

DR Interpro; IPR000859; CUB_domain.
 DR Interpro; IPR000421; FA58_C.
 DR Interpro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_RL_type_C; 1.
 DR PROSITE; PS0180; CUB; 1.
 DR DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 21.5%; Score 575.5; DB 4; Length 775;
 Best Local Similarity 24.5%; Pred. No. 2,4e-40;
 Matches 183; Conservative 84; Mismatches 181; Indels 299; Gaps 20;

QY 8 PSYALLFAVCAPLRLOAEELGDCGHIYVSODSGITMSKPNPTNYNYVCEIIVPK 67
 Db 48 PLFLILLVLLILLLEDAGAOQDCCGTYLGPESGITLSTNYPTYNSTVCEIIVK 107
 QY 68 GKRLILRLGDLNIE-SFTCASDYLLFSS---ATDOYGPYCG-SMAVPEKELRLNSMEY 121
 Db 108 GERVRIRKGFEDIDSDSCHFNRYLRINYNGIVSRTKGYGLQGNHSIESKNEITL 167
 QY 122 LFKSGSHISGRGFLTYASDHPDLITCLERGSHYFEERYSKFCPACGCDIADISGNT 181
 Db 168 LFKSGSHISGRGFLTYASDHPDLITCLERGSHYFEERYSKFCPACGCDIADISGNT 227
 QY 182 DGYRQSLCKAIIHAGIITDELGHIINLQSKGISHYEGLLANGVLSRHSLSERKFL 241
 Db 228 HGYRQSLCKAIIHAGIITDELGHIINLQSKGISHYEGLLANGVLSRHSLSERKFL 287
 QY 242 TT----- 243
 Db 288 KTSQCYGLMESGVADIPOITASSVLEMTDHTGOENSMKPKARKLPGPMAAFATDE 347
 QY 244 -----PGM----- 246
 Db 348 YQWLQIDLNKEKITGITTTGSTMVHEHNYYSAYRIIYSDGQKRTVYRBPVEQDKIFQ 407
 QY 247 ----- 246
 Db 408 GNDYHODVANNFLPPIIARFIRNPTQMOQKIAMKELLGQFIPKGRPKLTQPPPR 467
 QY 247 -----NIT-TV 251
 Db 468 NSNDLKNTTAPPKIARAKRFTQPLQPRSSNEPQTEQTASPDIRNTTVPNTYKDV 527
 QY 252 AIPSVI-FIALLLTGMIFAIC-----RKRKKGN---PYVSADAKTGCKQIYYPF 300
 Db 528 ALAAVLVPLVAVLTLLILLVCAHMRNRKKTETGYDLPY---WDRAGMKMGKQFL 583
 QY 301 -----ARHOSFTTISYNEKEMTKDLITSDMADYQPLMIGTVARKSGT 349
 Db 584 PAKAVDHEETPVARSSE--VNHSRPEVT---TVLQADAEATQAPLVGIVGTIQRST 638
 QY 350 FRPMDTIEEVRYNTEASG---YDCPRPGRHEVALPLTSHSEPEYAPIYERHLRLA 404
 Db 639 FRP-----EE---GKAGAGADLPYNSPCQEVYHAAVEPLITGPEYAPII----- 682
 QY 405 HTFSTGSGYVPG-PRPTHSHSGGFPATGATQVESYORPASPPVGGYDKP-AA 462
 Db 683 ---MDSGHPTTSYGQSTSTFKATYGNQPPLVGYNTLLSRDSSSAQAQYDTVPAGK 739
 QY 463 SFIDSRD-----PASQSOMTSGDDG 483
 Db 740 GGLPAPDELATYQYQPSQYQVSGAGRGS 766

RESULT 5
 Q88DX2 PRELIMINARY; PRT; 775 AA.
 AC Q88DX2;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

QY 254 PSYI--FIALLTGMGIFALC-----RRKKKG-NPIVSAQAOTGCKKQIK--YPPARHQ 304
 Db 525 AALVLPVLAVALTLLILVCAWHRNRKKAEGTDLPMDRAGWKKVKOLLPAKSVE 584
 QY 305 STEFTISYDNE-----KEMTOKLDTSDMDYOOPLMIGTVARKSTPRPHDPT 357
 Db 585 HEETPRVINSSEVSHLSPREVT---YVQADSAEYAOPLVGVGTLMHSTFRP----- 636
 QY 358 EEVRVNTAS-----GHYDCPRRGRHEVALPLTHSEPEVATPIVERHILRAHFTSTOSGY 413
 Db 637 EE---GKEASYADLPDYNAPVGEYHAHAYAPLPVTGGEVATPLV-----MDMSGH 683
 QY 414 RVVG-PRPTKHHSSGSGFPATGATOVESYQRPASPKPVGGYDKP-----AASFLD 466
 Db 684 STRASVGLPSTSTERTAGNOPALVGYNTLLSRDSCSSGOAQYDPEKGRKPAAPPELV 743
 QY 467 SRDPASOSQMTSGDD 482
 Db 744 YQVPOSTQEAAGACRD 759

RESULT 7

QY 091ZV3 PRELIMINARY; PRT; 769 AA.
 AC 091ZV3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 DE ESDN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR;
 RX PubMed=11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT *ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the longest Secretory Signal Sequence among Eukaryotes, is Up-regulated after Vascular Injury.*
 RT J. Biol. Chem. 276:34105-34114(2001).
 RL -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC EMBL: AF387548; AL30179.1; -;
 DR MGD: MGI:1920629; ESDN.
 DR InterPro: IPR000859; CUB_domaln.
 DR InterPro: IPR000421; FASB_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; transcription regulation.
 SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 21.0%; Score 560; DB 11; Length 769;
 Best Local Similarity 24.6%; Pred. No. 5e-39;
 Matches 191; Conservative 73; Mismatches 167; Indels 344; Gaps 23;

QY 4 GAGGSPV-----LALLFAVCAPL-----RL-----QAEELGDCGCHIV 37
 Db 15 GGGGPAAPATGAAALPSAGCCPLPGRRSSSRLLLLLLLLLLODAGGGGDCGHIVL 74
 QY 38 SDDSGMTSKNYRGTYPNTVCEKITTVPKGRLLIRLGDNLNE-SKTACSDYL-LPSS- 94
 Db 75 GPSSGLTSLINPHTYPSNVCEWEIRVTRGERIKIRKGFEDIEDSDYCHLNTKIFNGI 134
 QY 95 --ATDQGYPCG-SMAVPEKRLNSNEVTLFSGSHISGSGFLTYASSDHDHDLITCLE 151
 Db 135 GVSRTIIGKYGCGIAGLQMNOSIESKGESEVTVLFPASGTHAAGRGFLASTSYVDIKEDLLITCLD 194

QY 152 RGSHPFEKSKPCPACRCRIADIGSNTKDGNDTSLCKAAIHAGITDELGCHINL 211
 Db 195 TVSNFLPERFSKCPACCLPFAEISTGTHRGYDSSPLCMAGHAGVSNVLAGOISY 254
 QY 212 QSKGISHEGLANGVLSRGSUSEKRLP----- 241
 Db 255 ISKGTPESSLANNTVSTGYLSASLFTFTKSGCYGLMESGVADPQITASSALEMT 314
 QY 242 ----- 241
 Db 315 DHMGQENSWTEAKRLKPGPPMAAFATDEHQLDILNEKKITGIVTGTSMIEHSY 374
 QY 242 ----- 241
 Db 375 VSAFVLYSDGQKWTYRREPVDQDKIFQGNKYHKDYANNFLPPIAFIRVNVQW 434
 QY 242 ----- 241
 Db 435 OKIAMVELLCOFTLGRLPKLTTPPRNGNNTLNTTARPKLGKRAPKFTQVLOPRSN 494
 QY 242 -----TPGMNITVAIDSVI-----FIALLTGMGIFALC-----RRK 276
 Db 495 ELVPQPAETTTTDPDKNTVY-PSVTKDVALAVALVLAVALTLLILVCAWHRNRK 553
 QY 277 KK-GNPVSAADAKTGCWKQIKYF-----ARHSTFTISYDNEKEMTOKL 324
 Db 554 KKEGAVDLPHMPDRAGWKKMKOLLPAKSYDHEETPVRYSTSE-VSHLSAREVT---TV 608
 QY 325 ITSDMADYQOPLMIGTVARKSTPRMDTDEVRVNTASGH-----YDCPRRGRH 379
 Db 609 LOADSAEYAOPLVGVGTLMHSTFRP-----EE---GKEAGYADLPDYNAPVGEYHA 659
 QY 380 EYALPLTHSEPEVATPIVERHILRAHFTSTOSGYRVPGRPTKHHSSGSGFPA-----T 435
 Db 660 AYAEPLPVTGPEYATPLV-----MDMSG-----HPTASVGLSTSTFTK 698
 QY 436 GATOVESYQRPASPKPVGGYDKPRAASSFLDSRDPASOSQMTSGGDDGAPRNG 490
 Db 699 AGTQ-----PHLVGYTN-----TLISRTSCSSGOAQYDTPRGG 733

RESULT 8

QY 096NH2 PRELIMINARY; PRT; 155 AA.
 AC 096NH2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DE CDNA FLJ30900 fis, clone FEBRA2005752.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto T., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
 RT *NEO human cDNA sequencing project.*
 RL Submitted (Oct-2001) to the EMBL/Genbank/DBU databases.
 DR EMBL: AK055462; BAB70926.1; -;
 SQ SEQUENCE 155 AA; 16364 MW; 1C150371E07C2217 CRC64;

Query Match 18.8%; Score 503; DB 4; Length 155;
 Best Local Similarity 66.5%; Pred. No. 3.5e-35;
 Matches 103; Conservative 8; Mismatches 40; Indels 4; Gaps 3;

```

OY 353 MDTEEVNVTASGHDCPHRGRHEVALPLTHSEPEVATPIVERILLRAHFTSTSG 412
DB 1 MDTEAEAGVSTDAAGHDCPQAGRHEHYALPLAPPEPEVATPIYERHVALRAHFTSAQS 60
OY 413 YRVGPRPTHSHSSSGGFPATG-ATOVESYORASPKPVGGGCDKRAASFL--DSRD 469
DB 61 YRVGPRPTHSHSSSGGFPATG-ATOVESYORASPKPVGGGCDKRAASFL--DSRD 469
OY 470 PASOSQMT-SCGDGYSAPRNGALPLANTAMTALL 503
DB 121 PDSQKPTHGPTSDSYSAAPRDCPLPLNTAMTALL 155

RESULT 9
OY 014089 PRELIMINARY; PRT; 364 AA.
AC 014089;
DB 01-NOV-1996 (TREMBLrel. 01, Created)
DB 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.0 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shihata T.;
RT "unpublished.";
RL Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DB EMBL, D29810; BAI18909.1; -
DB InterPro: IPR000859; CUB_domain.
DB InterPro: IPR004043; FA58_C.
DB InterPro: IPR004043; LCCL_dom.
DB Pfam: PF00431; CUB; 1.
DB Pfam: PF00754; F5_P8_type_C; 1.
DB SMART: SM00042; CUB; 1.
DB PROSITE: PS01180; CUB; 1.
DR Hypothetical protein.
KW NON_TER
FT
SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Query Match 15.6%; Score 418; DB 4; Length 364;
Best Local Similarity 41.6%; Pred. No. 2,4e-27;
Matches 96; Conservative 34; Mismatches 79; Indels 22; Gaps 6;

OY 30 DCGHIVTSODSGTMTSKNRYPTNYCEKITVPRGKRLRLGLDNLTE-SKTCASD 88
DB 1 DCGHIVTSODSGTMTSKNRYPTNYCEKITVPRGKRLRLGLDNLTE-SKTCASD 88
OY 89 YLFSS---ATDQGYPCG-SMAVPEKLRLNSNEVTLFSGSHISGRGLTYASDH 143
DB 61 YLRIRNGVSRTEIGKYGGLQINHSIESKNEITLFLFSGIHVSGRGLTASYVDK 120
OY 144 PDLITCEHSGHYEERYSK-----CPAGCRDLADISGNTKDCYRPTSLCKRAI 195
DB 121 -----QRSVYLFGRHICIQEFNMLSVSTAPAGCLLPFAELISGTLPHGYRSPSCMAGV 173
OY 196 HAGITDELGHINLQSKGISHYEGGLANGLVSRHSL-SEKRFPLFTPG 245
DB 174 HAGVSTNLGGQISVIVSKIPYESSLIANNVTSVGHLLIQVFYFTKTSG 224

RESULT 10
OY 09CX06 PRELIMINARY; PRT; 414 AA.
AC 09CX06;
DB 01-JUN-2001 (TREMBLrel. 17, Created)
DB 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 2400001018RLX protein.
POCCE2 OR 2400001018RLX.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
DB MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull D., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DB EMBL, AK010249; BAB26794.1; -
DB InterPro: IPR000859; CUB_domain.
DB InterPro: IPR001134; Netrin_C.
DB Pfam: PF00431; CUB; 2.
DB Pfam: PF01759; NTR; 1.
DB SMART: SM00042; CUB; 2.
DB PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01180; CUB; 2.
SQ SEQUENCE 414 AA; 45408 MW; E51A29A352A2CF1D CRC64;

Query Match 6.7%; Score 178; DB 11; Length 414;
Best Local Similarity 23.4%; Pred. No. 1e-06;
Matches 82; Conservative 51; Mismatches 97; Indels 120; Gaps 18;

OY 1 MGTGAGPSVALLFPAVCAPRLQAEELG-DCGHIYTSODSGTMTSKNRYPTNYVC 59
DB 1 MGTGAGPSVALLFPAVCAPRLQAEELG-DCGHIYTSODSGTMTSKNRYPTNYVC 59
OY 60 EKITVPRGKRLRLGLDNLTE-SKTCASDYLFLSSADQ-----GPGGSAVAP 109
DB 59 TWKTIYPRGKRVVLFNFRIDLENOLCRDVF-----DVTNGHANGRIQIFGCTER-P 111
OY 110 KELRLNSNEVTLFSGSHISGRGLTYASD----- 141
DB 112 GSIVASGKMTVMQMSIDNMTNGSGFMATYSAADGKDRRCGRLEKPSGTFTPNMPD 171
OY 142 -DHPDLTC-----LEKSHYEEERYSKFCPAGCRDLADISGNTKDCYRPTSLCK 192
DB 172 RDYPGVGVVWHIAPKQNLIELKFEK-----DVE---RDNY-----CRYDYV 212
OY 193 AAIHAGITDELGHINLQSKGISHYEG----- 229
DB 213 AVENGGEVND-----AKRIGKYGDSPPVPIYERHVALRAHFTSAQS 262
OY 230 RHGSLSEKRFLETPGMNITVAIPSVIFIALLTGM-GIPAICRKRKK 278
DB 263 -HYKFRPKRFTTTTPTTTLTPV-----TGLKPTVALCOQKCR 302

RESULT 11
OY 08R4W6 PRELIMINARY; PRT; 414 AA.
AC 08R4W6;
DB 01-JUN-2002 (TREMBLrel. 21, Created)

```

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Procollagen COOH-terminal proteinase enhancer protein 2.
 GN PCOLCE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Steigltz B.M., Greenspan D.S.:
 RT Biochemical characterization and expression patterns of the
 RT Procollagen COOH-terminal proteinase enhancer protein 2.
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF352788; AAL83947.1;
 KM Collagen.
 SO SEQUENCE 414 AA; 45398 MW; 2A1A29A282A2CF1D CRC64;
 Query Match 6.7%; Score 178; DB 11; Length 414;
 Best local similarity 23.4%; Pred. No. 1e-06;
 Matches 82; Conservative 51; Mismatches 97; Indels 120; Gaps 18;
 QY 1 MGTGAGPSYLLAFNCALRLQAELG-DGCHYTSQDSGHTSKNTPTPTVTC 59
 DB 1 MG-GASACIPCLLALARRARPPQTPVFTGGILTG-ESGFISSGPGMPPMSKC 58
 QY 60 EKITYPCKRLILRLDLNIES-KTCASDYLLFSATDQY-----GPGCGSWAVP 109
 DB 59 TWKITPEGVVNLNPFIDLENDNLCRYDFV-----DYNGHANGQRIGRCGRTP-R 111
 QY 110 KEIRLNSNEVTVLFKSGHISGRGLTYVSS----- 141
 DB 112 GSIVASGNKMTYOMISDANAGSGFMATYSAAPDAGKGRGCRLEKPSGTFKTSNMPD 171
 QY 142 -DHPDLTC-----LENGSHFEERKSKFCPACGDIAGDISGNTKQGRDTSILCK---- 192
 DB 172 RDPVGTCTWHTIAPRNQILELKEFE-----DVE--RDNY-----CRDYY 212
 QY 193 AAIHAGITDELGHIMLQSGISHYEG----- 229
 DB 213 AVENGGEVND-----AKRIKCYGDSPPVPIYSENNELLIOFLSDLSLADFQIG 262
 QY 230 RRGSLSEKRFLETPPCNNITVAIPSVITALLITGM-GIFAICRRKKK 278
 DB 263 -HYKFPKPKKPTTPTTPTVTTLPV-----TTGLKPTVALCOQKCR 302
 RESULT 12
 Q8VBV4 PRELIMINARY; PRT; 694 AA.
 AC 08VBV4;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Hypothetical 77.4 kDa protein (Similar to complement component 1, s
 DE subcomponent).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC022123; AAH22123.1;
 DR EMBL; BC018319; AAH18319.1;

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00084; EGF; 1.
 DR Pfam: PF00089; Sush1; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00032; CCP; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00020; TRYPSIN_SPC; 1.
 DR PROSITE; PS00100; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01187; EGF_CA; UNKNOWN_1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
 KM Hypothetical protein.
 SO SEQUENCE 694 AA; 77415 MW; 06AA125D5E8227F5 CRC64;
 Query Match 6.4%; Score 172; DB 11; Length 694;
 Best local similarity 25.5%; Pred. No. 7.1e-06;
 Matches 54; Conservative 41; Mismatches 85; Indels 32; Gaps 9;
 QY 15 FAVCAP---LRQAELGDCGCHYTSQDSGHTSKNTPTPTVTCETITVPRKRL 71
 DB 161 FCSCPPEYFLHDDMRNGVNCSDVFTALIGELSSPPYPPYPPNSKCEYQIQEGSFQV 220
 QY 72 ILRL--GDLINE--SKTCASDYLLFSATDQYPCGS-WAVPKELRLNSNEVTVLFKS 125
 DB 221 VTMQREDFEVPADPSGNCPSDLTFASKNQCGPYGNGFPGLITRTQSNLTGIYFOT 280
 QY 126 GSHISGRGLTYVSSDHPDLITCL-----ERGSHYEERKSKFCPACGDIAG 174
 DB 281 DLMOQKQKMLRY---HGPDISCAKRTANSTWEPDKAYVFDVYKICVDSGFVEVG 336
 QY 175 DISGNT-----KDG-YRDSILCKAIIHAGI 199
 DB 337 HVSSTIYSTCQSDGWSNGLKQ-PVYCGI 367
 RESULT 13
 Q9UKZ9 PRELIMINARY; PRT; 415 AA.
 AC Q9UKZ9;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Procollagen C-terminal proteinase enhancer protein 2 (Procollagen
 DE C-proteinase enhancer protein 2).
 GN PCOLCE2 OR PCPE2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20334622; PubMed=10873381;
 RA Xu H., Acott T.S., Wirtz M.R.;
 RT Identification and Expression of a Novel Type I Procollagen C-
 RT Proteinase Enhancer Protein Gene from the Glaucoma Candidate Region on
 RT 3q21-q24.
 RL Genomics 66:264-273(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OSTEOARTHRITIC CARTILAGE;
 RX MEDLINE=21482651; PubMed=11597177;
 RA Kumar S., Connor J.R., Dodds R.A., Halsey W., Van Horn M., Mao J.,
 RA Sathe G.M., Mul P., Agarwal P., Badger A.M., Lee J.C., Gowen M.,

RA Lark M.W.:
 RT "Identification and initial characterization of 5000 expressed
 RT sequenced tags (ESTs) each from adult human normal and osteoarthritic
 RT cartilage cDNA libraries.";
 RL Osteoarthritis cartilage 9:641-653(2001).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC EMBL; AF098269; AAF04621.1; -;
 DR EMBL; AF098269; AAF04621.1; -;
 DR InterPro: IPR003439; ABC_transporth.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR001134; Netrin_C.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF01759; NTR; 1.
 DR SMART; SM00042; CUB; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS01180; CUB; 2.
 DR Collagen.
 KW SEQUENCE 415 AA; 45716 MW; B9A87FD9A87A5B CRC64;
 SQ
 Query Match 6.4%; Score 170.5; DB 4; Length 415;
 Best Local Similarity 23.3%; Pred. No. 4.5e-06;
 Matches 99; Conservative 67; Mismatches 152; Indels 107; Gaps 22;
 QY 4 GAGGSVALLEFAVCAPRLQ--AEELDGGCHITVSODSGMTSKNTPTGTPNTVCEK 61
 DB 3 GANAAAPICLLAAATQLSRQSPRPVPTGCGILTG--ESGFISGEFGPVPPMSKCTW 61
 QY 62 IIVPKGRLLRLDGLNIES-KTCASDYLLFSATDQY-----GPYCGSAVPEKE 111
 DB 62 KITVEGKVVVLFNFIDLESNLCRYFV-----DYRNGHANGORIGREGCTR--RGA 114
 QY 112 LRLNSNEVTVLFKSGSHISGRGFLITVAS-----D 142
 DB 115 LVSSGNKMMVOMISAPNAGNCFMAMFSAEPNERGDDYCGGLIDRPSGSEFTPMNPRD 174
 QY 143 HPDLTCTC--LENGSHYFEKYSKPCAG--CR-DIAGDISGNTKGYRPTSLCKAA 194
 DB 175 YPAGTCTWHYAPKNOILELKEFEKFDYERDNYCRDYAVNFGGEVNDARKIGYCDGS 234
 QY 195 IHAGITDELGHINLQSKGISHEGLANGVLSRHGSLSEKRLFTTPGNITTVAI 254
 DB 225 PPAPIVSR-----NELLIQFLSDL-SLADGFIGHY--IFRPKPLPTTQGPVT--TFP 285
 QY 225 SVIFAILLITGM-GIFALCRKRRK-----GNPYSAQAOKTCKKQIKYPPARHQSFEF 308
 DB 286 -----TTGLKPTVALCKCKRRTGLEGV-YCSSPFLVAG--TITTTTRGSLHA 334
 QY 309 TISYDN-----EKEWTKLDLITSDMADYQOPLMIGTGVARKSGTEFRPMDTD 356
 DB 335 TVSIINIKRGNLAIQAGKMSARLTVYCK-----QCPLL-----RGLNTYIIMGOV 382
 QY 357 TEVVR 361
 DB 383 GEDGR 387
 RESULT 14
 070542 PRELIMINARY; PRT; 694 AA.
 AC 070542;
 DT 01-AUG-1998 (Tremblurel. 07, Created)
 DT 01-AUG-1998 (Tremblurel. 07, Last sequence update)
 DT 01-JUN-2002 (Tremblurel. 21, Last annotation update)
 DE Sertine protease.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98192519; PubMed=9524231;
 RA Sakai H., Nakashima S., Yoshimura S., Nishimura Y., Sakai N.,
 RA Nozawa Y.;

RT "Molecular cloning of a cDNA encoding a serine protease homologous to
 RT complement C1s precursor from rat C6 glial cells and its expression
 RT during glial differentiation.";
 RL Gene 209:87-94(1998).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC EMBL; D88250; BAA25797.1; -;
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.193; -;
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00008; BGF; 1.
 DR Pfam; PF00084; Sush1; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00032; CCP; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00100; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS02440; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Calcium-binding; EGF-like domain; glycoprotein; Hydrolase; Protease;
 KW Repeat; Serine protease.
 KW SEQUENCE 694 AA; 77788 MW; CDC0ADFCEFFEC05 CRC64;
 SQ
 Query Match 6.3%; Score 169.5; DB 11; Length 694;
 Best Local Similarity 28.0%; Pred. No. 1.2e-05;
 Matches 52; Conservative 31; Mismatches 78; Indels 25; Gaps 7;
 QY 15 FAVCAP---LRLQAEELDGGCHITVSODSGMTSKNTPTGTPNTVCEKITVPEKRL 71
 DB 161 FCSCEPFEFLNDMDKTCGVNCGVFTALIGELASPNTPNPENSKCEYQIRLOEGFRL 220
 QY 72 IL--RLGDLNIE---SKTCADYLLFSATDQYGPYCGS--WAVPEKELNSNEVTVLFKS 125
 DB 221 VLTIRREDFEVPADSEGNCHDSLTFEAKNOQFGPYCGNGFPLTIRGQNTLDIYQIT 280
 QY 126 GSHISGRFLITVASDHPDLITC-----LENGSHYFEKYSKPCPAGCRDIAG 174
 DB 281 DLTGONKGMKRLRY---HGDPICPKETISANSIWEPEKAKYFVDYKVTICVDGEVEEG 336
 QY 175 DISGNT 180
 DB 337 NV-GST 341
 RESULT 15
 08R099 PRELIMINARY; PRT; 694 AA.
 AC 08R099;
 DT 01-JUN-2002 (Tremblurel. 21, Created)
 DT 01-JUN-2002 (Tremblurel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblurel. 21, Last annotation update)
 DE Similar to complement component 1, s subcomponent.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SALIVARY GLAND;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:36 ; Search time 11.5632 Seconds
(without alignments)
1279.897 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGSPVTLALFLAVCAP.....YSAPRNGIAPLQNTATALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfills1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164.5	6.2	415	4	US-09-032-523-2
2	143.5	5.4	923	3	US-08-936-135-6
3	142	5.3	986	4	US-08-872-757-4
4	140.5	5.3	788	1	US-08-572-225-1
5	139	5.2	1013	2	US-08-866-650-3
6	139	5.2	1013	2	US-09-021-287-3
7	139	5.2	1013	4	US-09-240-473-3
8	137	5.1	449	2	US-08-839-008-9
9	135	5.1	449	2	US-08-839-008-2
10	135	5.1	591	3	US-08-991-408-4
11	135	5.1	591	4	US-09-432-473-4
12	135	5.1	1013	2	US-08-866-650-5
13	135	5.1	1013	2	US-09-021-287-5
14	135	5.1	1013	3	US-08-991-408-2
15	135	5.1	1013	4	US-09-240-473-5
16	135	5.1	1013	4	US-09-432-473-2
17	133.5	5.0	468	4	US-09-032-523-8
18	132.5	5.0	909	3	US-08-936-135-18
19	132.5	5.0	926	3	US-08-936-135-20
20	131.5	4.9	922	4	US-09-116-473-4
21	131.5	4.9	925	4	US-09-116-473-2
22	130.5	4.9	1290	1	US-08-470-350B-2
23	129.5	4.8	901	3	US-08-936-135-22
24	129.5	4.8	906	3	US-08-936-135-24
25	129.5	4.8	909	3	US-08-936-135-8
26	129.5	4.8	909	3	US-08-936-135-10
27	129.5	4.8	914	3	US-08-936-135-12

28	129.5	4.8	926	3	US-08-936-135-14	Sequence 14, Appl
29	129.5	4.8	931	3	US-08-936-135-16	Sequence 16, Appl
30	126.5	4.7	855	2	US-09-027-337-2	Sequence 2, Appl
31	126.5	4.7	855	2	US-09-644-600-2	Sequence 2, Appl
32	126	4.7	401	2	US-08-839-008-5	Sequence 5, Appl
33	125.5	4.7	468	2	US-08-839-008-7	Sequence 7, Appl
34	125.5	4.7	666	4	US-09-341-587-1	Sequence 4, Appl
35	125.5	4.7	1785	4	US-09-341-587-3	Sequence 3, Appl
36	121	4.5	1019	1	US-08-296-014A-4	Sequence 4, Appl
37	121	4.5	1019	2	US-08-596-405-4	Sequence 4, Appl
38	121	4.5	1019	2	US-08-877-620-4	Sequence 4, Appl
39	121	4.5	1083	1	US-08-296-014A-2	Sequence 2, Appl
40	121	4.5	1083	2	US-08-596-405-2	Sequence 2, Appl
41	121	4.5	1083	2	US-08-877-620-2	Sequence 2, Appl
42	119	4.5	902	4	US-09-644-600-10	Sequence 10, Appl
43	118	4.4	102	4	US-09-374-135-7	Sequence 7, Appl
44	116.5	4.4	730	4	US-08-872-757-2	Sequence 2, Appl
45	116	4.3	1180	4	US-09-206-942-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-09-032-523-2
; Sequence 2, Application US/09032523

; Patent No. 6232454

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guebler, Karl

; APPLICANT: Baugh, Mariah

; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/032,523

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0479 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 415 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: RARRNOT02

; CLONE: 947429

; US-09-032-523-2

```

Query Match Similarity      6.2%  Score 164.5; DB 4; Length 415;
                          23.1%; Pred. No. 1.8e-08;
Matches 98; Conservative 67; Mismatches 153; Indels 107; Gaps 22.

    4 GAGGSPVALLAFVACAPRLQ--AELDGDGCGHYTSODSGMTSKNRYGTPNTVCEK 61
      || : || : || : || : || : || : || : || : || : || : || : ||
    Db 3 GANMAAPLCILLAAATQLSRQOSPFRPVTCGILLG--ESGFIGSGFGVYPPNSKCTW 61
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 62 IITVKGKRLIRLDNLIES-KTCASDYLLFSSATDQY-----GPGCNAVPEK 111
      |||| : || : || : || : || : || : || : || : || : || : || : ||
    Db 62 KTVTEGKRVYLNFRFIDLESNLCRYDFV-----DYNGHANGORIGRFGCFER-PCA 114
      |||| : || : || : || : || : || : || : || : || : || : || : ||
    Oy 112 LRLNENEVTLFSGSHISRGFLLTYASS-----D 142
      | : || : || : || : || : || : || : || : || : || : || : ||
    Db 115 LVSSGKMMVQVQIFDPAIRAGNGFMFSAEENEGDQYCGILLDRPSSFFTPWMPDRD 174
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 143 HPDLITC----LERGSHYFEERYSKFCPAG--CR-DIAGDISGNTKDGYPRTSLCKAA 194
      | : || : || : || : || : || : || : || : || : || : || : ||
    Db 175 YGAVTQCVWHIYAPRNQILTEKFEKFDYERNDYVCNRYDYAVNGGEVNDARIIGKYCGS 234
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 195 IHAGITBELGSHINLLQSKGISHYEGLIANGVLSRHGSLSKRPLFTTPGNNITVAIP 254
      || : || : || : || : || : || : || : || : || : || : || : ||
    Db 235 PPAPVYSEK-----NELLOFLSDS-SLVADFGIHY--IFRPKRLPTTEQPVTT-TEP 285
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 255 SVTFALLITGM-GIFAICRKRKK-----GNPYSAQAQKTCCKQIKYPPARHQSTFE 308
      || : || : || : || : || : || : || : || : || : || : || : ||
    Db 286 -----VITGLKPTVALCOQCRKRTGLENG-YCSSDVLVAG--VYTTTTRDGSLSHA 334
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 309 TISYDN-----EKEMTORLDLITSDMADYQOPLMIGTGVARKSGSTFRPMQTD 356
      || : || : || : || : || : || : || : || : || : || : || : ||
    Db 335 TVSIINIRKGNLAIQAGKKNMSARLTYVCK-----QCQLL-----RGLNITIMQCV 382
      || : || : || : || : || : || : || : || : || : || : || : ||
    Oy 357 TEVVR 361
      || : || : || : || : || : || : || : || : || : || : || : ||
    Db 383 GEDGR 387

RESULT 2
US-08-936-135-6
; Sequence 6, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: 0697-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 923 amino acids

```

```

1 TYPE: amino acid
2 STRANDEDNESS: single
3 TOPOLOGY: linear
4 MOLECULE TYPE: peptide
5 US-08-936-135-6
6
7 Query Match      5.4%; Score 143.5; DB 3; Length 923;
8 Best Local Similarity 19.3%; Pred. No. 1,le-05;
9 Matches 104; Conservative 80; Mismatches 199; Indels 157; Gaps 24;
10
11 QY 21 LRLQAEELGDCGGHIVTSODSGTMTSKNPGTYPNVTCETIITVPKGRLLILRLDLNI 80
12   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
13 Db 136 IYELEFKRNPESQNYTA-PTGVINSPGFPEKYPPCLCTEYIIIFAPKSEILLFESEFL 194
14   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 QY 81 ESKT-----CASDYLFESSATDDYGP---YCDSMAVPRKELRLNSNEVTLFRSGSHI 129
16   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17 Db 195 EDDSNRPFGMFCRYDRLEMTDMDFEPVEGPIGRKCQO-KTPGINRSSGVLNVFTDLSAI 253
18   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
19 QY 130 SGRGLLTVA---SSDHPLITLCERGSHYFEEKYSKEPCPACCRDIADISGNTRDYRD 186
20   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 Db 254 AKEGFSANYSVLQSSISEDF-----KCMALGMESGE----- 285
22   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
23 QY 187 TSLCLKAMIHGIT--DELGHINILLOSK-----GISHE-----GLL---- 223
24   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
25 Db 286 -----IHSQITPAASSQYGTMNSVERSRRLYPENGWTPGDSTSKEWIQVDLGLLRVT 337
26   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
27 QY 224 ---ANGVLISRHSLESEKRELFTTPGMNITTVAIPSVIF-----IALLNG-----MGI 268
28   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29 Db 338 AVGTGQALSKE---TKKKYYVTVRVSDISSNEDMISLESKNKAILIPGNTNPIDVVLGV 394
30   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
31 QY 269 FA-----ICRRKKKGKPNVVSADAOKTGCKWKQIYPPARH-----QSTEFPTISDN 314
32   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
33 Db 395 FKPPLITRFVRIRKPYSWETGISMRREYVGC-KITDYPGSGMLGNVSGLISDQITASNQA 453
34   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 QY 315 EKE-MTOKLDLITS-----DMADYO-Q-OPIMIGG----- 341
36   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 Db 454 DNNMKPENIRLVISTRTGVALPPSPHPYTNEMQLVDGDEKIYRGVITIOGKRREKVFNR 513
38   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
39 QY 342 ---TWARKGSTFRPMDDTEEVARNVTEASGHYDCPHRPGRHVALPLTHSEPEXATPI- 396
40   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
41 Db 514 KKRIAYSNNGSDWMKTINDSKRAKASFEGENNNYDPTELTFPSLTRIRIYPERATHSG 573
42   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43 QY 397 --VEHHLLRAHFQSOSQGYRVGPRPETH-----HSHSGCGPPATGATOVSYORP 446
44   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
45 Db 574 ICLRPELLGCEYEAFETAGPTTPNGNPRVHCDDQDANCHSGTDDE-QLTGCTVLATERP 632
46   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47
48 RESULT 3
49 US-08-872-757-4
50 Sequence 4, Application US/08872757
51 Patent No. 6258584
52 GENERAL INFORMATION:
53 APPLICANT: Prockop, Darwin J.
54 APPLICANT: Hojima, Yoshio
55 APPLICANT: Li, Shi-wu
56 APPLICANT: Steron, Aleksander
57 TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
58 TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF
59 NUMBER OF SEQUENCES: 6
60 CORRESPONDENCE ADDRESS:
61 ADDRESSEE: Pennie & Edmunds
62 STREET: 1155 Avenue of the Americas
63 CITY: New York
64 STATE: New York
65 COUNTRY: U.S.A.
66 ZIP: 10036-2711
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS
71 SOFTWARE: PatentIn Release #1.0, Version #1.30
72 CURRENT APPLICATION DATA:
73 APPLICATION NUMBER: US/08/872.757

```


RESULT 4
 US-08-572-225-1
 : Sequence 1, Application US/08572225
 : Patent No. 5807981
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Prockop, Darwin J.
 : APPLICANT: Hojlma, Yoshio
 : APPLICANT: Li, Shi-Mu
 : APPLICANT: Steron, Alexander
 : APPLICANT: Brenner, Mitch
 : TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND ITS USE FOR
 : TITLE OF INVENTION: DRUG DEVELOPMENT FOR THE TREATMENT OF DISEASE
 : NUMBER OF SEQUENCES: 24
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Pennile & Edmonds
 : STREET: 1155 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/572.225

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/866,650
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Bersco, Bennett, J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296, 93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-866-650-3

```

Query Match 5.2%; Score 139; DB 2; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYRYVCEKIITVPGKRLRLRGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPHRTTLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YPCGSAVNAVKELRLNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFQATHSTECGRLKAESKPRDL 900
QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSESGS 927

RESULT 6

US-09-021-287-3
Sequence 3, Application US/09021287
Patent No. 5981717
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tollold-Like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,287
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/866,650
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296.93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-021-287-3

Query Match 5.2%; Score 139; DB 2; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYRYVCEKIITVPGKRLRLRGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPHRTTLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YPCGSAVNAVKELRLNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFQATHSTECGRLKAESKPRDL 900

QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSESGS 927

RESULT 7

US-09-240-473-3
Sequence 3, Application US/09240473
Patent No. 6297011
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tollold-Like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/240,473
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296.93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-240-473-3

Query Match 5.2%; Score 139; DB 4; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYRYVCEKIITVPGKRLRLRGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPHRTTLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YPCGSAVNAVKELRLNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFQATHSTECGRLKAESKPRDL 900
QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSESGS 927

RESULT 8

US-08-839-008-9
Sequence 9, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R

APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-9

Query Match 5.1%; Score 137; DB 2; Length 449;

Best Local Similarity 24.8%; Pred. No. 1.6e-05; Matches 62; Conservative 32; Mismatches 98; Indels 58; Gaps 11;

QY 5 AGPSVALLFAVCAPRLQAEELGDS-----CGHIVTSQDSGTMTSKNYPGTYPTV 58
DB 4 AATASLGLPLTACALLPFAOGQTPNTRPVFLCGGVKG-ESGYVASEGFPNLYPPNKE 62
QY 59 CEKIIIVPKGRLL--RLGLDNIESTKASDYL-LFS---SATDGYPCGSNAVPKEL 112
DB 63 CIIITIVPEQGVSLSFVVDLEH--PACHYDALEVPAGSGTSGQRLGRFCGFR--PAPL 120
QY 113 RLNSNEVTVLFKSGSHSGFLITYA-----SSDH 143
DB 121 VAPGNQVTLMTDEGTGGRGFLMTSGRATSGEHQFCGRLEKAOGTLLTPNWPESDY 180
QY 144 PDLITCLERSHYE-----EKYSKFCPACGR-DIAGDISGNTKGYRDTSLCK 192
DB 181 PPGISC---SMHIIAPPDVIATLFEKFDLEPDTYCRDVSVPNGAVSDSRLGRFCG 237
QY 193 AAIIHAGITTD 202
DB 238 DAVPGSISSE 247

RESULT 9
US-08-839-008-2
Sequence 2, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A

APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-2

Query Match 5.1%; Score 135; DB 2; Length 449;

Best Local Similarity 24.8%; Pred. No. 2.6e-05; Matches 62; Conservative 31; Mismatches 99; Indels 58; Gaps 11;

QY 5 AGPSVALLFAVCAPRLQAEELGDS-----CGHIVTSQDSGTMTSKNYPGTYPTV 58
DB 4 AATASLGLPLTACALLPFAOGQTPNTRPVFLCGGVKG-ESGYVASEGFPNLYPPNKE 62
QY 59 CEKIIIVPKGRLL--RLGLDNIESTKASDYL-LFS---SATDGYPCGSNAVPKEL 112
DB 63 CIIITIVPEQGVSLSFVVDLEH--PACHYDALEVPAGSGTSGQRLGRFCGFR--PAPL 120
QY 113 RLNSNEVTVLFKSGSHSGFLITYA-----SSDH 143
DB 121 VAPGNQVTLMTDEGTGGRGFLMTSGRATSGEHQFCGRLEKAOGTLLTPNWPESDY 180
QY 144 PDLITCLERSHYE-----EKYSKFCPACGR-DIAGDISGNTKGYRDTSLCK 192
DB 181 PPGISC---SMHIIAPPDVIATLFEKFDLEPDTYCRDVSVPNGAVSDSRLGRFCG 237
QY 193 AAIIHAGITTD 202
DB 238 DAVPGSISSE 247

RESULT 10
US-08-991-408-4
Sequence 4, Application US/08991408
Patent No. 6008017
GENERAL INFORMATION:
APPLICANT: ARLETH, ANTHONY J.
APPLICANT: WILLETTTE, ROBERT N.
APPLICANT: ELSHOUBAGY, NABIL A.

US-08-866-650-5

Query Match	5.1%;	Score 135;	DB 2;	Length 1013;
Best Local Similarity	23.1%;	Pred. NO. 0.0001;		
Matches 42;	Conservative 31;	Mismatches 67;	Indels 42;	Gaps 8;

[illegible]

RESULT 13
US-09-021

```

Sequence 5, Application US/05021287
Patent No. 5981717
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tollold-like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Charles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,287
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/866,650
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296,93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-021-287-5

```

```
Query Match      5.1% Score 135; DB 2; Length 1013;  
Best Local Similarity 23.1%; Pred. No. 0.0001;  
Matches 42; Conservative 31; Mismatches 67; Indels 42; Gaps 8;  
  
QY    32 CGHIYSDSGMTMSKNPTGTYPPNTVQEKITTVKGRRLIIRLDDLNLT-ESKTCASDYL 90  
|| : : : : : | ||| || | : : | : : : : | : : : : | || :
```

32 CGHIVTSQDSGIMTSKNYPGTYPRNTYCEKITVYKGRLLIRLGDINI-ESKTCASDYL 90

Db 349 CGETL-QESNGNLSSPGFPNGYPSYTHCIWRVSVTPGEKIVLNF¹TTMDLYKSSLCWYDI 407

Dy 91 LFSATQY-----GPYCSMAVPEKLRLNSENVTLPKSGSHISGRGLFTYASDH 143
Dy 408 ---EVVDGYWPKSPLLARPCSD-KLPELVISTDSRMILERRSSSNWGKGPAAYEA--- 460
Dy 144 PDLTYTCERGSHYPEEKYSKPCPAGCRDIAGDI-SGMTKDGYRDT-----SLCKAIIHA 197
Dy 461 -----ICGEIKNNEQIQDSPNYPPDYRPMEKCEWKITVSESRYHV 500

OY 198 GI 199
Db 501 GL 502

RESULT 14

US-08-991-408-2
Sequence 2, Application US/08991408
Patent No. 6008017
GENERAL INFORMATION:
APPLICANT: ARLETH, ANTHONY J.
APPLICANT: WILLETTTE, ROBERT N.
APPLICANT: ELSHOURBAGY, NABIL A.
APPLICANT: LI, XIAOTONG
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLOID-LIKE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATTNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,408
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/034,471
FILING DATE: 02-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50038
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-408-2

Query Match	5.1%	Score 135;	DB 3;	length 1013;
Best Local Similarity	23.1%;	Pred. No. 0.0001;		
Matches 42;	Conservative 31;	Mismatches 67;	Indels 42;	Gaps 8

QY 32 CGHIVTSQDSSTSKNVPGRYPVATYCKEIKIIPVKCRILLIRGDLNI-ESKTCASDYL 90
|| : : : || : | : | : | : : : : : : : : : : : :
Db 349 CGEHL-OESNKLSPPFPMGYPIRCIMRVSTPEKEILVFETWIDLLKSSLCWYDI 4070
| :
QY 91 LFSATQY-----GPYGGMNVPEELRNINSEVTVLEKSGSHIGRGFLITYASDH 143

Search completed: May 15, 2003, 13:15:23
Job time : 16.5632 secs

Db 408 ---EVRDGYWRKSPILGRPCGD-KLPEVLTSTDSNMWIEFRSSNMWVGKFAVYEA--- 460
QY 144 PDLITCLERGSHYFEKYSKFCPACGRDIAGDI-SGNTKDGTRD-----SLICKAAIHA 197
Db 461 -----ICGGEIRKNEGQIQSPNTPDDYRPMKECVWKITVSESYHV 500
QY 198 GI 199
Db 501 GI 502

RESULT 15

US-09-240-473-5
; Sequence 5, Application US/09240473
; Patent No. 6297011
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Takahara, Kazuhiko
; APPLICANT: Hoffman, Guy G
; TITLE OF INVENTION: Mammalian tolloid-like Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/240.473
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berson, Bennett J
; REGISTRATION NUMBER: 37094
; REFERENCE/DOCKET NUMBER: 960296.93839
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1013 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-240-473-5

Query Match 5.1%; Score 135; DB 4; Length 1013;
Best Local Similarity 23.1%; Pred. No. 0.0001;
Matches 42; Conservative 31; Mismatches 67; Indels 42; Gaps 8;

QY 32 CGHIVTSDSGTMTSKNPGYVNTVCEKITVPRKRLILRLGLDINT-ESKTCASDYL 90
Db 349 CGEIL-QESNGNLSPGPNPGSYTHCIMRWVSYPGERIVLNTFTMDYKSSLCWYDI 407
QY 91 LESSATDOY-----GYYCGSMVAPKELRLNSNEVTLFKSGSHISGRGFLITVAAASDH 143
Db 408 ---EVRDGYWRKSPILGRPCGD-KLPEVLTSTDSNMWIEFRSSNMWVGKFAVYEA--- 460
QY 144 PDLITCLERGSHYFEKYSKFCPACGRDIAGDI-SGNTKDGTRD-----SLICKAAIHA 197
Db 461 -----ICGGEIRKNEGQIQSPNTPDDYRPMKECVWKITVSESYHV 500
QY 198 GI 199
Db 501 GI 502

QY	241	FTTPGMMITVADPSVFIALLLGMGIFAIICRRKKKKGNPYVSADQKIGCMKQIKRYP	300
Db	241	FTTGMMITVADPSVFIALLLGMGIFAIICRRKKKKGNPYVSADQKIGCMKQIKRYP	3000
QY	301	ARHOSTEFTTISYDNKEKEMTKLDLITSDMADYQOPLMIGTVARKSGTERPMDTTEEV	360
Db	301	ARHOSTEFTTISYDNKEKEMTKLDLITSDMADYQOPLMIGTVARKSGTERPMDTTEEV	3600
QY	361	RVNTEASGHHDCPRRPGRHEHVALPLTSEPEVATPIYERHLLRAHNETSGRGVPPRP	420
Db	361	RVNTEASGHHDCPRRPGRHEHVALPLTSEPEVATPIYERHLLRAHNETSGRGVPPRP	4200
QY	421	TKHSHSSSGGPPATGATQVESYORPASPRVGGGYDKPAASFLDRDPAASQSOMTSGG	480
Db	421	TKHSHSSSGGPPATGATQVESYORPASPRVGGGYDKPAASFLDRDPAASQSOMTSGG	4800
QY	481	DDGYSAPRNGIAPLNOTAMRALL 503	
Db	481	DDGYSAPRNGIAPLNOTAMRALL 503	

RESULT 2
 US-09-823-038A-51
 Sequence 51, Application US/09823038A
 Patent No. US20020058335A1
 GENERAL INFORMATION:
 APPLICANT: Strachan, Iorna
 APPLICANT: Sleeman, Matthew
 APPLICANT: Abernethy, Nevian
 APPLICANT: Onrust, Rene
 APPLICANT: Kumble, Anand
 APPLICANT: Murison, Greg
 TITLE OF INVENTION: Compositions Isolated From Stromal Cells
 TITLE OF INVENTION: and Methods For Their Use
 FILE REFERENCE: 11000.1037/c3
 CURRENT APPLICATION NUMBER: US/09/823,038A
 CURRENT FILING DATE: 2001-07-09
 NUMBER OF SEQ ID NOS: 61
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 51
 LENGTH: 503
 TYPE: PRT
 ORGANISM: Mouse
 US-09-823-038A-51

Query Match	99.68;	Score 2659;	DB 10;	Length 503;
Best Local Similarity	99.68;	Pred. No. 4.1e-209;		
Matches 501; Conservative	1;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	1	MGTAGGGSVTLALFVACAPLRLOAEELGBOCCGIVTSOSGTMTSKNYGGTYPNTVYCE	60
Db	1	MGTAGGGSVTLALFVACAPLRLOAEELGBOCCGIVTSOSGTMTSKNYGGTYPNTVYCE	60
Qy	61	KITVPKPKRRLILAGDLINBKTCSADYLLFSSATQYGPYCGSMAVPKELRLNSVYT	120
Db	61	KITVPKPKRRLILAGDLINBKTCSADYLLFSSATQYGPYCGSMAVPKELRLNSVYT	120
Qy	121	VLFKSGSHISGRGELLYTVAASDHPDLITTCLESGSHYFEERYSKRCPAGCADIAGDISGNT	180
Db	121	VLFKSGSHISGRGELLYTVAASDHPDLITTCLESGSHYFEERYSKRCPAGCADIAGDISGNT	180
Qy	181	KDGTRDTSILCKAAIHAGIITDELGGHINLQSKGISHYEGLLANGVLSRHGSISEKRL	240
Db	181	KDGTRDTSILCKAAIHAGIITDELGGHINLQSKGISHYEGLLANGVLSRHGSISEKRL	240
Qy	241	FTTPGGMNTTVAIPSVIFIALLLGMSGIFALCRKRKKKGNPYVSADAKRTGCKWQIKYPF	300
Db	241	FTTPGGMNTTVAIPSVIFIALLLGMSGIFALCRKRKKKGNPYVSADAKRTGCKWQIKYPF	300
Qy	301	ARHSTETTSYDNKEKEMTKQLDILTSDMADYQOPLMIGITVARKSSTRPMDTDEEV	360
Db	301	ARHSTETTSYDNKEKEMTKQLDILTSDMADYQOPLMIGITVARKSSTRPMDTDEEV	360

Qy	361	RVNTEASGHHYCCPRPGGHEHALPLTSEPEYATPIVERHLLRHNTESTQSGRVPEPR	420
Db	361	RVNTEASGHHYCCPRPGGHEHALPLTSEPEYATPIVERHLLRHNTESTQSGRVPEPR	420
Qy	421	THKSHSGGPPATGATQVSYSTORPASPKVGGGYDKPAASSFLDSRDSASQMTSGG	480
Db	421	THKSHSGGPPATGATQVSYSTORPASPKVGGGYDKPAASSFLDSRDSASQMTSGG	480
Qy	481	DDGYSAPRNGLAFLNQTAMTALL 503	
Db	481	DDGYSAPRNGLAFLNQTAMTALL 503	

RESULT 3
 US-10-003-132-6
 : Sequence 6, Application US/10003132
 : Publication NO. US20020192750A1
 : GENERAL INFORMATION:
 : APPLICANT: Fox, Brian A.
 : APPLICANT: Gao, Zeren
 : APPLICANT: Shoemaker, Kimberly E.
 : TITLE OF INVENTION: NEUROBILIN HOMOLOG ZCUBS5
 : FILE REFERENCE: 00-62
 : CURRENT APPLICATION NUMBER: US/10/003,132
 : CURRENT FILING DATE: 2001-11-15
 : PRIOR APPLICATION NUMBER: US 60/249,004
 : PRIOR FILING DATE: 2000-11-15
 : NUMBER OF SEQ ID NOS: 19
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 6
 : LENGTH: 458
 : TYPE: PRT
 : ORGANISM: Mus musculus
 US-10-003-132-6

Query Match	89.7%	Score 2395.5	DB 9	Length 458
Best Local Similarity	91.1%	Pred. NO. 1.2e-187		
Matches 458; Conservative	0	Mismatches	0	Indels 45; Gaps 1.

Qy	1	MGTAGGVSVALLEFVACAPRLRLAEELGADCCGHITVTSOSGTTMTSKNYGTPNTVYCE	60
Dd	1	MGTAGGVSVALLEFVACAPRLRLAEELGADCCGHITVTSOSGTTMTSKNYGTPNTVYCE	60
Qy	61	KITVPKCKRLILRLGDLNIESKTCASDYLLFSSATDOYGYCGSWAVPELRLNSNEVT	120
Dd	61	KITVPKCKRLILRLGDLNIESKTCASDYLLFSSATDOY-----	99
Qy	121	VLFKSGHISGRGFLLYASSDHPDLITCLERSGHYEEBKSKFPCPACRDIADISGNT	180
Dd	100	-----DLITCLERSGHYEEBKSKFPCPACRDIADISGNT	135
Qy	181	KDGYRDTSLCKKAIHAGITDELGSHNLLNOSGISHYEELANGLVLSHRGSLSEKRL	240
Dd	136	KDGYRDTSLCKKAIHAGITDELGSHNLLNOSGISHYEELANGLVLSHRGSLSEKRL	195
Qy	241	FTTPGMNTTVAISVIFIALLLTGMGIFAICRRKKRKKGNPYVSADAKTGCKWKQIYYP	300
Dd	196	FTTPGMNTTVAISVIFIALLLTGMGIFAICRRKKRKKGNPYVSADAKTGCKWKQIYYP	255
Qy	301	ARHOSTEETIYDNEKEKTKLIDITSMDADYQOPLMIGTGVAKCSTTRPMDTTEEV	360
Dd	256	ARHOSTEETIYDNEKEKTKLIDITSMDADYQOPLMIGTGVAKCSTTRPMDTTEEV	315
Qy	361	RVNTEASGHYCCPHRPGHEVALLTSEPEYAPPIYERHLLRHTSTOSGYVPEPRP	420
Dd	316	RVNTEASGHYCCPHRPGHEVALLTSEPEYAPPIYERHLLRHTSTOSGYVPEPRP	375
Qy	421	THKHSHGSGPPATGATOVESYORPASPKVGVGGYKRPAASSFILDRDPAOSOSOMTSGG	480
Dd	376	THKHSHGSGPPATGATOVESYORPASPKVGVGGYKRPAASSFILDRDPAOSOSOMTSGG	435
Qy	481	DDGYSAPRNGLAPLNQTAMTALL 503	


```

DB          436 DDGYSAPRNGLAPLNQNTAMTALL 458

RESULT 4
US-10-003-132-2
; Sequence 2, Application US/10003332
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROBILIN HOMOLOG ZCDB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

```

Query Match	72.1%	Score 1924.5;	DB 9;	Length 715;
Best Local Similarity	55.8%;	Pred. No. 7.3e-149;		
Matches 392;	Conservative 34;	Mismatches 73;	Indels 203;	Gaps 4

0Y	5	AGDSVTLALPACAPLQAEELGDCGHVTSQDSGTMSSKNYPCTYPIVYCEKTI	64
Db	14	AAGRLALALLAVAPLRLQAEELGDCGHLVYQDSGTMSSKNYPCTYPIVYCEKTI	73
0Y	65	VPKGRLLRLRGDLNIESKTCASDYLLPSSATDQYGYCGSWAVPKELRLNSNEVYLFK	124
Db	74	VPKGRLLRLRGDLNIESQTCASDYLLFTSSDDQYGYCGSMVPEKELLNTSEVYRFE	133
0Y	125	SGSHSGGFLTLTVAASSHPDLICLERSHYPEEKSKFCPCACCRDADISGNTDGY	184
Db	134	SGSHSGGFLTLTVAASSHPDLICLERASHYLTETSKFCPCACCRVADISGNMVDGY	193
0Y	185	RDTSLLCKAAIHAGITDELGHNILNLSKGISHYEGILLANGVLSRRHSSEKRFLE	241
Db	194	RDTSLLCKAAIHAGIADIELGQISVLRKGISHYEGILLANGVLSRDSLSDKRFLETSN	253
0Y	242	-----241-----	
Db	254	GCSNLSFEPDQIRASSWSQVNESGDVHMSPGQARLQDQGSWASGDSNNHKPREW	313
0Y	242	-----241-----	
Db	314	LEIDLGKKKTTGTRTGSSTOSNPFYVKSIVMFKNNKSKMKTGKIYNNKEEVQDGS	373
0Y	242	-----241-----	
Db	374	NFRDPVONNEPPIVARYRVVQTMHQRLAKYELIGCOITGNDSLVNRKTSQTSYS	433
0Y	242	-----TPGKNITTVAILPSYFIALLTGMGFALICORKKKKGNPYSA	285
Db	434	TKKEDETTRPPEEJEUSTGINTITTVAILPLVLVLVAFAGGJFPAARKKRKKGSPGSA	493
0Y	286	DAQTGCKWQKQJTEPPARHOSQTEFTISYDNKEKMQKDLITSDMAVDQOPLMGTGVAR	345
Db	494	EAQGTDCMKQIKQJFPARRHQSAPFTISYDNKEKMQKDLITSDMAVDQOPLMGTGVYTR	553
0Y	346	KGSTFRPMDDTTEVRYNTEASGHYDCPHRPGHEHYALPLTHSEPEYATPIVERHLLRAH	405
Db	554	KGSTFRPMDDTDAEBAAGVSTDAAGHYDCPORAGRHEHYALPLAPEPEYATPIVERHYLRAH	613
0Y	406	TFSTOSGVRVGPPTKHKHSSSGGPPRATG-ATQVESYORPASPCKVGGGYYOKPAASSP	464
Db	614	TFSMOSGVRVGPQGRHKHSLSSGGFSVAVGACQDQYQPHSAQPADKQIDTPKAVSA	673

```

QY      465 L-DSRDPAISQSQMT-SGDDDEGSAPRNGLPANQAMTALL 503
      : : | | | | | | | | | | | | | | | | | | | |
Db      674 LATESGHPSQKPTNPGTSDSYSAKDCLPLNQMAMTALL 715

```

```

RESULT 5
US-09-759-130B-73
: Sequence 73, Application US/09759130B
: Publication No. US20030022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D
: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirsch, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodearl, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: FILE REFERENCE: MP100-535ONMIM
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 73
: LENGTH: 715
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-759-130B-73

```

[illegible]

```
|||||
Db 194 RPTSLCKAAIHAGIITADELGGQISVLQKRGISREGLANGVLSRDGSLSKRFLTSN 253
Oy 242 ----- 241
Db 254 GCSRLSEFPDQIIRASSWSQVNSGQVHMSPGQARLQDQSPSWASGSSNNHKPREM 313
Oy 242 ----- 241
Db 314 LEIDGERKKITGIRTTGTSQSNFNFYKSFVNFKNNSKWKTKYGVNNEKYFGNS 373
Oy 242 ----- 241
Db 374 NFRDVOYNNFPIYARVYVPPQTHORIALKVELIGCQITQGNDLSLWRTSSTSYS 433
Oy 242 -----TPGNNITVAIPSVIFALLLTGMGIFAICRRKKKKPYSA 285
Db 434 TKKEDEITRPISEETSGINITVAIPLVLLVLFVAGMGIFAFRKKKKSPYSGA 493
Oy 286 DAQKTGCKWQIKYPPARHOSSTFTSYDNEKEMTKLDTISDMADYQOPLMIGTVAR 345
Db 494 EAKQTDCKWQIKYPPARHOSSTFTSYDNEKEMTKLDTISDMADYQOPLMIGTVAR 553
Oy 346 KGFTRPMDTDEEYRVNTEASGHDCPHRGRHEVALPLTSEPEYATPIYERHLLRAH 405
Db 554 KGFTRPMDTDEEYRVNTEASGHDCPHRGRHEVALPLTSEPEYATPIYERHLLRAH 613
Oy 406 TFSSTGCVRVPPRPHKSHSSGSGFPATG-ATQVESYQRPASPKPYGCGYDKPAASSF 464
Db 614 TFSAGSGRVPPQPGHSHSSGSGFPATGAGDQYQRHSAQPADRGYDRKAVSA 673
Oy 465 L--DSRDPASQSQMT--SGGDDGYSAPRNGLAPLNOTAMTALL 503
Db 674 LATESGHPDSQKPPHPPGTSYSAPRDCLTPLNOTAMTALL 715

RESULT 6
US-09-759-130B-75
; Sequence 75, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarty, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirsch, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: MP100-5350M1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
```

```
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-75

Query Match 69.6%; Score 1859.5; DB 9; Length 681;
Best Local Similarity 55.4%; Pred. No. 1,4e-143;
Matches 377; Conservative 33; Mismatches 68; Indels 203; Gaps 4;

Oy 26 EELGCGCHITVSQDSGMTSKNYPTIPNTYCEKIIITVPRKRLILRLGDLNIESTC 85
Db 1 EELGCGCHITVSQDSGMTSKNYPTIPNTYCEKIIITVPRKRLILRLGDLNIESTC 60
Oy 86 ASDYLLFSATDQYVPGGSMVAPKELRLNSNEVTVLPKSGSHISGRGFLTYASSDHPD 145
Db 61 ASDYLLFTSSDQYVPGGSMVAPKELRLNSNEVTVLPKSGSHISGRGFLTYASSDHPD 120
Oy 146 LITCERSHYEERKSKPCACRDIAGDISGNKRGYRDTSLCKAIAHAGIITDEIG 205
Db 121 LITCERSHYEERKSKPCACRDIAGDISGNKRGYRDTSLCKAIAHAGIITDEIG 180
Oy 206 GHINLOSIGSHYEGGLANGYLSRHSSEKRTLF----- 241
Db 181 GQISVLRKGISRYEGILANGVLSRHSSEKRTLF----- 240
Oy 242 ----- 241
Db 241 SYNESGDOYHMSPGQARLQDQSPSWASGSSNNHKPREMLEIDLEKKITGIRTTGTSQ 300
Oy 242 ----- 241
Db 301 SNFNFYKSFVNFKNNSKWKTKYGVNNEKYFQNSNFRDPYONNFIPIYARVYRV 360
Oy 242 -----TTGCM 246
Db 361 VPQTHORIALKVELIGCQITQGNDLSLWRTSOSTSVTKREDETTRPIPSEETSGI 420
Oy 247 NITVAIPSVIFALLLTGMGIFAICRRKKKKGNVYSADQKTCMKQIKYPPARHOS 306
Db 421 NITVAIPLVLLVLFVAGMGIFAFRKKKKSPYSGAEMQKTCMKQIKYPPARHOSA 480
Oy 307 EFTISYDNEKEMTKLDTISDMADYQOPLMIGTVARKSGTFRPMDTDEEYRVNTEA 366
Db 481 EFTISYDNEKEMTKLDTISDMADYQOPLMIGTVARKSGTFRPMDTDEEYRVNTEA 540
Oy 367 SGHYCCPHRPGHHEVALPLTSEPEYATPIYERHLLRAHTEFGSGYVPPGGRPHKSH 426
Db 541 GGHYCCPHRPGHHEVALPLTSEPEYATPIYERHLLRAHTEFGSGYVPPGGRPHKSH 600
Oy 427 SSGGFPPATG-ATQVESYQRPASPKPYGCGYDKPAASSFL--DSRDPASQSQMT--SGGDD 482
Db 601 SSGGFSPVAGVADQGDYQRPASQPADRGYDRPAAVSALESHPDSQKPPHTPIGSD 660
Oy 483 GYSAPRNGLAPLNOTAMTALL 503
Db 661 GYSAPRNGLAPLNOTAMTALL 681

RESULT 7
US-09-759-130B-76
; Sequence 76, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
```

Query Match	37.3%	Score 995;	DB 9;	Length 421;
Best Local Similarity	84.1%;	Pred. No. 3.3e-73;		
Matches 185; Conservative	18;	Mismatches 17;	Indels 0;	Gaps 0

Db 181 GQISVLQKGISRYEGILANGVLSRDGSLSDKRLFTSNG 220

; APPLICANT: McCarthy, Sean A .

Query Match	33.48;	Score 893;	DB 9;	Length 235;
Best Local Similarity	74.98;	Pred. No. 3.2e-65;		
Matches 176;	Conservative 11;	Mismatches 44;	Indels 4;	Gaps 3;

Db 181 ADRGYDRPKAVSALATESGHPDSQKPTTHPGTSDSYAPRDLTPLNQNTAMTALL 235

1. TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN

FILE REFERENCE: PB0169
CURRENT APPLICATION NUMBER: US/10/060, 830
CURRENT FILING DATE: 2002-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/864, 761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/325, 062
PRIOR FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 1123
SOFTWARE: Neomica Sequence Listing Engine
SEQ ID NO 3
LENGTH: 729
TYPE: PRT
ORGANISM: Homo sapiens
US-10-060-830-3

Query Match 21.5%; Score 575.5; DB 9; Length 729;
Best Local Similarity 24.5%; Pred. No. 1.4e-38;
Matches 183; Conservative 84; Mismatches 181; Indels 299; Gaps 20;

QY 8 PSVALLEAVCAPRLQAEELGDCGGCHVTSDSGTMTSKNPGTYPVNYCEKITYPK 67
DB 2 PLFLLLVLLVLLLEDDAGAOOGDGGCHTVLSPESGTLTINPQYTPNSVCEWEIRYK 61
QY 68 GKRLRLGLDNLIE-SKTCASDYLLFS-ATDQYPCYG-SWAVPKELRLNSNEYTV 121
DB 62 GERVAIKGDDIEDSDSCHNYLTNYGIVSKREIKRYCGLGOMHNSISKNETIL 121
QY 122 LFKSGSHISGFGFLTYVASDHPDLITCLERSHYFEERYSKFCPAGCRDIAGDISGNTK 181
DB 122 LFMGSHVSGGFLASVVIDKODLITCLDPAASNLEPEFSKCPAGCLLPFAELISGTP 181
QY 182 DGYRDTSLCAAIHAGITTDGLGCHINLQSKGSHYEGLLANGLVLSRHGSLSKRFL 241
DB 182 HGYSRDSPLCMAGVAGVSNLGGQISVLSKGIPIYESSLIANNVTSVGHLSLFTF 241
QY 242 TT----- 243
DB 242 KTSQCYGLMGESVIADPQITASSVLEWTDHTGOENSMKPKKARKPGPPMAAFATDE 301
QY 244 -----PGM----- 246
DB 302 YQMLQIDLNKKKIKITGTTGTVWEHNYVSAYRIYSDQKQWTVYREGVEQDKIFQ 361
QY 247 ----- 246
DB 362 GNKDYHODVRNNLPPIIARIRVNPQMOQKIAKMKELLCQGITPKGRPKLTQPPPR 421
QY 247 -----NIT-TV 251
DB 422 NSNDLKNTPAKTAKGAPFTQPLQPRSSNEPPAQTEQITASPDIINTVTPVNTADV 481
QY 252 AIPSVI-FIALLLTGMGIFAIC-----RRKKKGN-EVVSADAKQTCGKWKQKYP 300
DB 482 ALAAVLVPLVAVLTTLLILVCAWHNRNKKKTKGTGTYDLP-----WDRAGWMMGKQFL 537
QY 301 -----ARHSTFETSYDNEKEMTOKLILITSDMADYQOPLMIGITVAKKSGT 349
DB 538 PAKAVDHEETPVRYSSSE-VNHLSPREVT-TVLQADSAYEAPLIVGIGVTLHQRST 592
QY 350 FRPMDTDEEYRVMTSEAGH-----YDCPHRGRHEVALPLTHSRPEVATPIVERHLLRA 404

DB 593 KRP-----EKEGAYDLDPYNSGQEVYHAIVAEPLITGPEYATPII----- 636
QY 405 HTFSTOSGYRVPG-PRPTHKSHSSGEPFPATGATQVESYORPASPKEVGGYDKP-AAS 462
DB 637 ---MMSGHPTTSVGQPSPTSFKATGNQPPPLVGYVNTLLSTDBCSSAQAOYDPRKAGK 693
QY 463 SFLDSRD-----PASOSQMTSGGDDG 483
DB 694 PELPAPDELVYQVPQSTQEVSGAGRDG 720

RESULT 10

US-10-060-830-1114
Sequence 1114, Application US/10060830
Publication No. US20030032154A1
GENERAL INFORMATION:
APPLICANT: Gu, Yizhong
APPLICANT: Nguyen, Cung-Tuong
TITLE OF INVENTION: HUMAN LCCL DOMAIN CONTAINING PROTEIN
FILE REFERENCE: PB0169
CURRENT APPLICATION NUMBER: US/10/060, 830
CURRENT FILING DATE: 2002-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/864, 761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/325, 062
PRIOR FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 1123
SOFTWARE: Neomica Sequence Listing Engine
SEQ ID NO 1114
LENGTH: 653
TYPE: PRT
ORGANISM: Homo sapiens
US-10-060-830-1114

Query Match 15.6%; Score 417; DB 9; Length 653;
Best Local Similarity 22.3%; Pred. No. 1.1e-25;
Matches 147; Conservative 70; Mismatches 149; Indels 294; Gaps 18;

QY 90 LLESSATDQYPCYG-SWAVPKELRLNSNEYTVLFKSGSHISGRGFLTYVASDHPDLIT 148
DB 13 LLEDDAGAQQKRYCGLGOMHNSIESKNGNEITLLFMGSHVSGGFLASVVIDKODLIT 72
QY 149 CLERSHYFEERYSKFCPAGCRDIAGDISGNTKDCYRPTSLCAAIHAGITTDGLGHI 208
DB 73 CLDTASNLEPEFSKYKCPAGCLLPFAELISGTPHGYRDSPLCMAGVAGVSNLGGQI 132
QY 209 NLQSKGSHYEGLLANGLVLSRHGSLSEKRELFET 243
DB 133 SVVLSKGIPIYESSLIANNVTSVGHLSLFTFKTSGCYGLMGESVIADPQITASSVL 192
QY 244 ----- 243
DB 193 EWTDHTGOENSMKPKKARKRKPPMAAFATDEYQMLQIDLNKKKIKITGTTGTVWEH 252
QY 244 -----PGM----- 246
DB 253 NYVSAYRIYSDQKQWTVYREGVEQDKIFQGNKDYHODVRNNLPPIIARIRVNP 312
QY 247 ----- 246
DB 313 QMOQKIAKMKELLCQGITPKGRPKLTQPPPRNSNDLKNTPAKTAKGAPFTQPLQ 372

```

OY 247 -----NIT-TVAIPSVI--FIALLITGICPAIC--- 272
DB 373 PRSSNEPQTEQOTASPDIRNTVTPTNKDVALAVALPVLMVTLTILIVCAMW 432
OY 273 --RKRKKGN--PYVSADAKTCCKKQIKYPP-----ARHQSREPTISDNKR 316
DB 433 RNRKKTEGYDLPY--WDRAKMKMGKQFLPARAVDHEETPVRSSE--VNHLSPR 486
OY 317 EMTOKDLTSDMADYQOPLMIGTIVARKSTFRPMQDTEEVRYNTEASGH-----YD 371
DB 487 EYT---TVLOADSAEYAOPLVGIYVGLMORSTFKP-----EE-----GKAGTADLDPTN 534
OY 372 CPHRPGRHEVALPLTHSEPEYATPIVERHLLRAHTESTOSGYRVPG--PRPTKHSHSSG 430
DB 535 SPQOEYHYHVAEPLPTIGPEYATPII-----MDMSGHPTTSVGPSTSTFRATGN 584
OY 431 FPPATATQVESQORAPSRPVGGYDKP--AASSFLDSRP-----PASQOMTSGGDDG 483
DB 585 QPPLVGTNTLTLSKRDSCSSAQAOYDTPKAGRPGLPAPDELVYOVPOSTOEVSGAGRDG 644

```

RESULT 11

```

US-09-874-198-2
Sequence 2, Application US/09874198
Patent No. US20020082208A1
GENERAL INFORMATION:
APPLICANT: Jensenius, Jens Chr.
APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
FILE REFERENCE: 09011-002002
CURRENT APPLICATION NUMBER: US/09/874,198
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 686
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-198-2

```

```

Query Match 6.6%; Score 177.5; DB 10; Length 686;
Best Local Similarity 24.4%; Pred. No. 4.3e-06;
Matches 52; Conservative 39; Mismatches 97; Indels 25; Gaps 7;

```

```

OY 32 CGHIVTSQDSGTMTSKNYPTGYPNYVCEKIIIVPKGRLLT---RLGDLNIESKT-CAS 87
DB 184 CSQOVFTQSGELSPREYPRPKLSSCTYSISLEGFVYLDVFSFDETHPELTCY 243
OY 88 DYLLFSASATDQYGYCGSNAVPRKELRLNSNEVTLF---KSGSHISGRGFLTYASSDHP 144
DB 244 DFLKIQTDREHPCFGK--TLPHRIETKSTVYITVTDESNGH---TGKRIHTYTAOP 299
OY 145 DLITCLERESH-----YEEKYKFCPCAGCDIAGDISGNT-----KDGYPDTSL 190
DB 300 CYPMAAPPNGHSPVQAKYILKDSFSIFCETGYELLQGHILPKSFTAVCKDKDSMDRMP 359
OY 191 CKRAIHAGITDELGHINILQSKGISHYEGL 223
DB 360 ACSIVDCGPPDDLPSGRVEXYITGPGVITYKAVI 392

```

```

RESULT 12
US-09-874-238-2
Sequence 2, Application US/09874238
Patent No. US20020082209A1
GENERAL INFORMATION:
APPLICANT: Jensenius, Jens Chr.

```

```

APPLICANT: Thiel, Steffen
TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
FILE REFERENCE: 09011-002003
CURRENT APPLICATION NUMBER: US/09/874,238
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/054,218
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 60/042,678
PRIOR FILING DATE: 1997-04-03
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 686
TYPE: PRT
ORGANISM: Homo sapiens
US-09-874-238-2

```

```

Query Match 6.6%; Score 177.5; DB 10; Length 686;
Best Local Similarity 24.4%; Pred. No. 4.3e-06;
Matches 52; Conservative 39; Mismatches 97; Indels 25; Gaps 7;

```

```

OY 32 CGHIVTSQDSGTMTSKNYPTGYPNYVCEKIIIVPKGRLLT---RLGDLNIESKT-CAS 87
DB 184 CSQOVFTQSGELSPREYPRPKLSSCTYSISLEGFVYLDVFSFDETHPELTCY 243
OY 88 DYLLFSASATDQYGYCGSNAVPRKELRLNSNEVTLF---KSGSHISGRGFLTYASSDHP 144
DB 244 DFLKIQTDREHPCFGK--TLPHRIETKSTVYITVTDESNGH---TGKRIHTYTAOP 299
OY 145 DLITCLERESH-----YEEKYKFCPCAGCDIAGDISGNT-----KDGYPDTSL 190
DB 300 CYPMAAPPNGHSPVQAKYILKDSFSIFCETGYELLQGHILPKSFTAVCKDKDSMDRMP 359
OY 191 CKRAIHAGITDELGHINILQSKGISHYEGL 223
DB 360 ACSIVDCGPPDDLPSGRVEXYITGPGVITYKAVI 392

```

```

RESULT 13
US-09-864-761-45371
Sequence 45371, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30

```

```

PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00660
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 45371
LENGTH: 75
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC013497.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
OTHER INFORMATION: EST_HUMAN HIT: A1565996.1, EVALUE 2.00e-40
OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09
US-09-864-761-45371

Query Match 6.4%; Score 170.5; DB 10; Length 75;
Best Local Similarity 54.8%; Pred. No. 7.2e-07;
Matches 34; Conservative 9; Mismatches 18; Indels 1; Gaps:

QY 30 DCGCHYTSDDSGTMTSKMYPCGYTNYVYCEKITVPRGRRLIRLDDNIE-SKTCASD 88
Db 1 DCGCHYVLGSGSTLTFSINYPTYPNSTVYCEWEIRVKGGEHVRIRIKFDPIEDSDSCHFN 60
QY 89 YL 90
||
Db 61 YL 62

RESULT 14
US-09-905-291A-104
Sequence 104, Application US/0905291A
Patent No. US20020160374A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Iyar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann

```

```

: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 104
: LENGTH: 415
: TYPE: PRT
: ORGANISM: Homo Sapien
US-09-902-853-104

Query Match          6.4%; Score 170.5; DB 9; Length 415;
Best Local Similarity 23.3%; Pred No. 8e-06;
Matches 99; Conservative 67; Mismatches 152; Indels 107; Gaps 22

   4 GAGGSPVALLAFVACAPLRLQ--AELADGCGHIVTSODSGTMTSKNYPGYPNVTVCK 61
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db  3 GANMAAPCLILAAATQLSROOSPFRPVYTCGILTG-ESGFIGSGFGVYPPNSKCTW 61

   62 ITTVKGRKRLRLGDNIET-KTCASDYLLFSSATDQ-----GPYCGMAVPKE 111
   ||| : || : || : || : || : || : || : || : || : || : || : ||
Db  62 KTTVEGKRVVILNFEFIDESDNLCRYDQV-----DYNGHANGORIGFCGTER-PCA 114

   112 LRLNENVTYLFKSGSHISRGFLLTYASS-----D 142
   | : || : || : || : || : || : || : || : || : || : || : || : ||
Db  115 LVSSGNKMWVOMISDANTRAGNGFMAMFSAEPNERGDOYCGLLDRPSSGFKTPNNMPDRD 174

   143 HPDLITC-----LERGSHYEEERYSKFPCAG---CR-DIAGDISGNTKDGVRDTSLLCKAA 194
   | : || : || : || : || : || : || : || : || : || : || : || : ||
Db  175 YPAGVTCWVHIYAPKNQDLEKFEKFDVDRDNYCRDYAVYNGGEVNDARIGKYGDS 234

   195 IHAGITDELGSHINLQSKGISHEGLIANGVLSRHGSLSEKRFLETTPGANITTVAIR 254
   || : || : || : || : || : || : || : || : || : || : || : ||
Db  235 PPAPIVSER-----NELLIQFLSDL-SLTADGFIGHY--IFRPKKLPTTTEQPVTT-TPP 285

   255 SVIFALLLTGM-GIFATCRKRRKK-----GNPYVSADAKTGCKWQIKYPPARHSTEF 308
   || : || : || : || : || : || : || : || : || : || : || : ||
Db  286 -----VTTGLKPTVALCQOKCRRTGLEGN-YCSSDVLVAG--TVTTTTRDGSLSLA 334

   309 TISYDN-----EKEMTQKLDLITSDMADYQOPLMIGTGTVARKGSTFRPMQTD 356
   | : || : || : || : || : || : || : || : || : || : || : || : ||
Db  335 TVSIINIRKEGMLAIQOAGKMSARLITVCK-----QCPLL-----RGLNITIMGOY 382

   357 TEEVR 361
   | : |
Db  363 GEDGR 367

Search completed: May 15, 2003, 13:21:36
Job time : 20.9319 secs

```

THIS PAGE BLANK (USPTO)